

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.

2771

500.38010X00

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

RECEIVED

FEB 17 2000

Applicant(s): TAKAHASHI, et al.

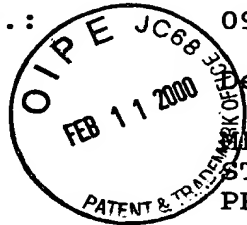
Serial No.: 09/461,192

Filed: December 15, 1999

Title: METHOD AND SYSTEM FOR MANAGEMENT OF
STRUCTURED DOCUMENT AND MEDIUM HAVING
PROCESSING PROGRAM THEREFOR

Group:

GROUP 2700



K. Ward
3/23/00
#3/Pr. Priority Paper

LETTER CLAIMING RIGHT OF PRIORITY

Honorable Commissioner of
Patents and Trademarks
Washington, D.C. 20231

February 11, 2000

Sir:

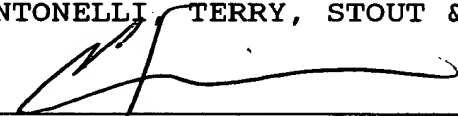
Under the provisions of 35 USC 119 and 37 CFR 1.55, the
applicant(s) hereby claim(s) the right of priority based on:

Japanese Patent Application No. 10-360110
Filed: December 18, 1998

A certified copy of said Japanese Patent Application is
attached.

Respectfully submitted,

ANTONELLI, TERRY, STOUT & KRAUS, LLP



Carl I. Brundidge
Registration No. 29,621

CIB/ssr
Attachment

日 本 国 特 許 庁
PATENT OFFICE
JAPANESE GOVERNMENT

RECEIVED

FEB 17 2000

GROUP 2700

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日
Date of Application:

1998年12月18日

出 願 番 号
Application Number:

平成10年特許願第360110号

出 願 人
Applicant(s):

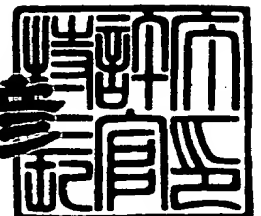
株式会社日立製作所



1999年12月10日

特許庁長官
Commissioner,
Patent Office

近 藤 隆 彦



出証番号 出証特平11-3086364

【書類名】 特許願

【整理番号】 K98009881

【提出日】 平成10年12月18日

【あて先】 特許庁長官 殿

【国際特許分類】 G06F 17/00

【発明の名称】 構造化文書管理方法及びその実施装置並びにその処理プログラムを記録した媒体

【請求項の数】 20

【発明者】

【住所又は居所】 神奈川県川崎市幸区鹿島田 890 番地 株式会社日立製作所 システム開発本部内

【氏名】 高橋 亨

【発明者】

【住所又は居所】 神奈川県横浜市戸塚区戸塚町 5030 番地 株式会社日立製作所 ソフトウェア事業部内

【氏名】 山崎 紀之

【発明者】

【住所又は居所】 神奈川県川崎市幸区鹿島田 890 番地 株式会社日立製作所 システム開発本部内

【氏名】 里 佳史

【特許出願人】

【識別番号】 000005108

【氏名又は名称】 株式会社日立製作所

【代理人】

【識別番号】 100083552

【弁理士】

【氏名又は名称】 秋田 収喜

【電話番号】 03-3893-6221

【手数料の表示】

【予納台帳番号】 014579

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 構造化文書管理方法及びその実施装置並びにその処理プログラムを記録した媒体

【特許請求の範囲】

【請求項 1】 文書記述言語を用いて記述された構造化文書を管理する構造化文書管理方法において、

登録済み文書の実体構造及び論理構造を示す対応関係データから登録済み文書を構成する任意のファイルを更新対象として選択して更新対象ファイルの内容データを更新する内容データ更新ステップと、更新後の更新対象ファイルの実体構造及び論理構造を示す部分的な対応関係データを生成する部分的対応関係データ生成ステップと、前記部分的な対応関係データによって登録済み文書の対応関係データを更新する対応関係データ更新ステップとを有することを特徴とする構造化文書管理方法。

【請求項 2】 前記内容データ更新ステップは、外部プログラムによって編集された更新後の内容データによって更新対象ファイルの内容データを置換することを特徴とする請求項 1 に記載された構造化文書管理方法。

【請求項 3】 前記部分的対応関係データ生成ステップは、前記更新対象ファイルが文書ファイルまたはテキストファイルである場合に更新後の更新対象ファイルを解析して部分的な対応関係データを生成し、そのいずれでもない場合に更新後の内容データそのものを設定した部分的な対応関係データを生成することを特徴とする請求項 1 または請求項 2 のいずれかに記載された構造化文書管理方法。

【請求項 4】 前記更新対象ファイルのファイル構成及び解析済みインスタンスデータを示す構文解析結果データを生成する文書構文解析ステップと、前記更新対象ファイルを構成する各ファイルと解析済みインスタンスデータにより得られる論理構造部分とを対応付けて部分的な対応関係データを生成する対応関係データ抽出ステップとを有することを特徴とする請求項 3 に記載された構造化文書管理方法。

【請求項5】 登録済み文書の対応関係データ中で前記更新対象ファイルに対応する部分範囲を求める対応部分範囲決定ステップを有し、

前記対応関係データ更新ステップは、前記求めた部分範囲の対応関係データを前記更新対象ファイルの部分的な対応関係データによって置換して登録済み文書の対応関係データを更新することを特徴とする請求項1乃至請求項4のいずれか1項に記載された構造化文書管理方法。

【請求項6】 前記更新対象ファイルにロックを設定して当該ファイルへの編集処理を排他的に行うことを特徴とする請求項1乃至請求項5のいずれか1項に記載された構造化文書管理方法。

【請求項7】 文書記述言語を用いて記述された構造化文書を管理する構造化文書管理方法において、

登録済み文書の実体構造及び論理構造を示す対応関係データから登録済み文書の任意の論理構造を更新対象として選択して編集する論理構造編集ステップと、前記編集した論理構造の範囲内にあるファイルの実体構造及び論理構造を示す部分的な対応関係データを文字列化して記述した編集結果データを生成する編集結果データ生成ステップと、前記生成した編集結果データによって前記ファイルの内容データを置換する対応内容データ更新ステップと、前記部分的な対応関係データによって登録済み文書の対応関係データを更新する対応関係データ更新ステップとを有することを特徴とする構造化文書管理方法。

【請求項8】 登録済み文書の対応関係データ中で前記ファイルに対応する部分範囲を求める更新対象検出ステップを有し、

前記対応関係データ更新ステップは、前記求めた部分範囲の対応関係データを前記ファイルの部分的な対応関係データによって置換して登録済み文書の対応関係データを更新することを特徴とする請求項7に記載された構造化文書管理方法。

【請求項9】 前記更新対象の論理構造の範囲内にあるファイルにロックを設定して当該ファイルへの編集処理を排他的に行うことを特徴とする請求項7または請求項8のいずれかに記載された構造化文書管理方法。

【請求項10】 登録済み文書の持つファイル構成に対応して設定されたノード集合の属性に応じてノード集合毎に更新対象へのアクセス制御を行うことを特徴とする請求項1乃至請求項9のいずれか1項に記載された構造化文書管理方法。

【請求項11】 登録済み文書を構成するファイル単位で編集対象を取り出して更新処理を行うことを特徴とする請求項1乃至請求項10のいずれか1項に記載された構造化文書管理方法。

【請求項12】 複数の編集対象を別々に取り出して編集し、更新後の編集対象を再登録することにより構造化文書の編集を並列に行うことを特徴とする請求項1乃至請求項11のいずれか1項に記載された構造化文書管理方法。

【請求項13】 更新後の編集対象を構成するファイルに電子署名を付加することを特徴とする請求項1乃至請求項12のいずれか1項に記載された構造化文書管理方法。

【請求項14】 登録対象文書のファイル構成及び解析済みインスタンスデータを示す構文解析結果データを生成する文書構文解析ステップと、登録対象文書を構成する各ファイルと解析済みインスタンスデータにより得られる論理構造部分とを対応付けて登録対象文書の実体構造及び論理構造を示す対応関係データを生成する対応関係データ抽出ステップとを有することを特徴とする請求項1乃至請求項13のいずれか1項に記載された構造化文書管理方法。

【請求項15】 登録対象文書の持つファイル構成に対応したノード集合を設定し、対応するファイルの持つ属性を継承してその属性を各ノード集合毎に付与することを特徴とする請求項14に記載された構造化文書管理方法。

【請求項16】 登録対象文書を構成するファイルに付加された電子署名を当該ファイルの内容データと共に登録することを特徴とする請求項14または請求項15のいずれかに記載された構造化文書管理方法。

【請求項17】 文書記述言語を用いて記述された構造化文書を管理する構造化文書管理装置において、

登録済み文書の実体構造及び論理構造を示す対応関係データから登録済み文書を構成する任意のファイルを更新対象として選択して更新対象ファイルの内容デ

ータを更新する内容データ更新処理部と、更新後の更新対象ファイルの実体構造及び論理構造を示す部分的な対応関係データを生成する部分的対応関係データ生成処理部と、前記部分的な対応関係データによって登録済み文書の対応関係データを更新する対応関係データ更新処理部とを備えることを特徴とする構造化文書管理装置。

【請求項 18】 文書記述言語を用いて記述された構造化文書を管理する構造化文書管理装置において、

登録済み文書の実体構造及び論理構造を示す対応関係データから登録済み文書の任意の論理構造を更新対象として選択して編集する論理構造編集処理部と、前記編集した論理構造の範囲内にあるファイルの実体構造及び論理構造を示す部分的な対応関係データを文字列化して記述した編集結果データを生成する編集結果データ生成処理部と、前記生成した編集結果データによって前記ファイルの内容データを置換する対応内容データ更新処理部と、前記部分的な対応関係データによって登録済み文書の対応関係データを更新する対応関係データ更新処理部とを備えることを特徴とする構造化文書管理装置。

【請求項 19】 文書記述言語を用いて記述された構造化文書を管理する構造化文書管理装置としてコンピュータを機能させる為のプログラムを記録した媒体において、

登録済み文書の実体構造及び論理構造を示す対応関係データから登録済み文書を構成する任意のファイルを更新対象として選択して更新対象ファイルの内容データを更新する内容データ更新処理部と、更新後の更新対象ファイルの実体構造及び論理構造を示す部分的な対応関係データを生成する部分的対応関係データ生成処理部と、前記部分的な対応関係データによって登録済み文書の対応関係データを更新する対応関係データ更新処理部としてコンピュータを機能させる為のプログラムを記録したことを特徴とする媒体。

【請求項 20】 文書記述言語を用いて記述された構造化文書を管理する構造化文書管理装置としてコンピュータを機能させる為のプログラムを記録した媒体において、

登録済み文書の実体構造及び論理構造を示す対応関係データから登録済み文書

の任意の論理構造を更新対象として選択して編集する論理構造編集処理部と、前記編集した論理構造の範囲内にあるファイルの実体構造及び論理構造を示す部分的な対応関係データを文字列化して記述した編集結果データを生成する編集結果データ生成処理部と、前記生成した編集結果データによって前記ファイルの内容データを置換する対応内容データ更新処理部と、前記部分的な対応関係データによって登録済み文書の対応関係データを更新する対応関係データ更新処理部としてコンピュータを機能させる為のプログラムを記録したことを特徴とする媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明はコンピュータ装置を用いた電子化文書データを管理する構造化文書管理システムに関し、特に文書の論理的構造とそのエンティティ構成とを対応付けて管理する構造化文書管理システムに適用して有効な技術に関するものである。

【0002】

【従来の技術】

情報化社会の本格的進展に伴い、ワードプロセッサ、パーソナルコンピュータ等を用いて作成された電子化文書データが爆発的な勢いで増加しつつある。単に雑多な文書をコンピュータ上で作成して表示や印刷に用いるだけでなく、組織体の業務上重要な役割を果たす文書（ミッションクリティカル文書）を電子的に作成し、それらの文書を取り扱う業務全体を電子化することによって、業務効率の大幅な向上を実現することが望まれている。

【0003】

この様なミッションクリティカル文書は、単に表示・印刷のみを目的とするものではなく、文書内部からのデータの抽出、派生文書の作成や表示体裁を変えての再利用等、さまざまな処理の対象となる為、多くの場合SGML(Standard Generalized Markup Language)、XML(eXtensible Markup Language)、HTML(Hyper Text Markup Language)等の文書記述言語を用いて、容易に機械処理可能な構造化文書データとして作成される。従って、これらのミッションクリティカル文書を扱う文書管理システムは、構造化文書を格納・管理し、構造情報の参照、文書内容

の編集・改訂、バージョン履歴の管理等を行う機能を備える必要がある。

【0004】

前記文書記述言語によって記述された構造化文書は、実体構造と論理構造との二重の構造を持つ。ここで論理構造とは、文書を構成する個々の論理的な要素（章、節、段落その他）の並び順や相互の包含関係からなる構造であり、前記文書記述言語では、その要素の種別（要素型）を示す文字列である「タグ」を各論理要素の先頭と末尾に配置することによって文書の論理構造を表現する。また実体構造とは、文書データを「エンティティ」と呼ぶ単位に分割し、内容テキスト中に他のエンティティへの参照を記述することによって表現される構造であり、エンティティ間の呼び出し関係からなる。

【0005】

一般に、文書が長大になり、単一の連続したテキストとして文書を維持するのが不適切になった場合、文書を複数のエンティティに分割し、各エンティティ毎に担当者を割り当てて個別に執筆・編集する等の管理が行なわれる。また、図表データの様に各種の処理プログラム固有のデータ形式で記述され、内容文字列やタグからなるテキストと混在させることができないデータも、それぞれ別のエンティティとして扱われる。文書が複数のエンティティから構成される場合、通常は各エンティティをそれぞれ1個のファイルに格納する。なお、構造化文書中で要素を識別する為に用いられるタグ、エンティティへの参照を記述するエンティティ参照、及びコメント等を記述した文字列を総称してマークと呼ぶ。

【0006】

図15は従来の文書記述言語SGMLによって記述された構造化文書の一例を示す図である。1個のSGML文書の実体構造は、文書本体に対応する1個のエンティティ（文書エンティティ）と、ここから直接間接に参照される0個以上の外部エンティティから構成される。

外部エンティティには、文書エンティティと同様にSGMLテキストとして記述されるテキストエンティティと、図表等の非SGMLデータエンティティとがある（本明細書では、文書エンティティと外部エンティティとを総称して単にエンティティと呼ぶ）。図15に例示する文書の場合、文書エンティティ101からテキス

トエンティティ 102 及び 103 並びに非 SGML データエンティティ 104 が参照されている。またテキストエンティティ 102 中にもテキストエンティティ 103 への参照がある。その結果、テキストエンティティ 103 は文書エンティティ 101 とテキストエンティティ 102 の両方から参照され、両者の間で共有されている。

【0007】

図 16 は従来の図 15 の SGML 文書を解析して得られる論理構造の例を示す図である。図 16 に示す様に SGML 文書の持つ論理構造は、要素と内容データをノードとする木構造として表現することができる。

図 16 において楕円形のノードが要素、矩形のノードが内容データ（解析済み文字列データまたは非 SGML データ）を表している。また図中で点線で囲った範囲はそれぞれノードの集合を表している。ここでノード集合 201 は前記の文書エンティティ 101 に対応しており、ノード集合 202 はテキストエンティティ 102、ノード集合 203 及び 204 はテキストエンティティ 103、ノード集合 205 はデータエンティティ 104 に対応している。エンティティとノード集合との対応関係が示す通り、エンティティ参照によって呼び出した被参照エンティティの内容をその呼び出し位置に展開し、結果として得られる展開済みテキストを構文解析することにより、SGML 文書の論理構造が得られる。

【0008】

従来、文書データはファイルに格納され、ファイルシステム上で管理されてきた。特開平 9-223054 号公報に示される発明（以後、これを公知例 1 と呼ぶ）は、複数のファイルからなるファイル集合を単位として、記憶手段からのファイルの読み出し及び記憶手段へのファイルの格納、またファイル集合を単位とするバージョン管理を実現している。公知例 1 は本来コンピュータプログラムを構成するファイル群の管理を目的とした発明であるが、そのファイル管理機能は大規模な文書を構成するファイル群を管理する目的にも使用することができる。しかし、公知例 1 の発明は、構造を持たないデータとして個々のファイルを扱うので、文書管理に適用した場合に文書の構造を意識した操作を行うことはできない。

【0009】

これに対して構造化文書を直接管理できる従来技術として、例えば特開平8-44718号公報に示された発明（以後、これを公知例2と呼ぶ）がある。公知例2に示される文書処理装置は、登録された文書の構造を解析し、章、節等の論理要素をノードとする木構造データを生成してこれを格納・管理する手段を備えている。また文書構造の一部を構成する部分木を複数文書間で共有する手段や、部分木に対して論理要素の追加/削除/更新等が行われた場合に、そのバージョンを管理する手段を備えている。

【0010】

【発明が解決しようとする課題】

前記公知例1に示す従来技術では、文書を構成する個々のファイル及びそれらのファイルの集合としての文書の状態を管理することはできるが、文書の持つ論理的構造を認識し、構造に沿った操作を行う手段を持たないので、構造化文書を扱う文書管理システムとしては適していない。これに対して前記公知例2に示す従来技術では、登録された文書の持つ論理的構造を表現する木構造データを直接格納・管理するので、格納されている構造化文書データに対して構造情報の参照や編集、部分構造（部分木）単位の共有制御、部分構造のバージョンの管理等が可能となる。しかし、SGMLやXML等の文書記述言語を用いて記述された構造化文書を管理しようとした場合、前記公知例2に示す従来技術によっても以下に説明する様な課題が解決されずに残る。

【0011】

すなわち前記公知例2に示す従来技術は、文書の論理構造を格納・管理する手段は備えているものの、実体構造と論理構造とを対応付けて管理する手段や、木構造形式で表現された対応関係データと文書記述言語を用いてテキスト形式で記述されたエンティティデータとの間で随時相互変換を行う手段は備えていない。前記公知例2に示す従来技術を適用した文書管理システムにSGML文書を登録した場合、登録前にそのSGML文書が持っていたエンティティ構成に関する情報は失われてしまう。

【0012】

この為、前記公知例 2 に示す従来技術を適用した文書管理システムに複数のエンティティから構成される SGML 文書を登録して論理構造データを生成する場合、登録対象文書の持つエンティティ構成に対応したノード集合を設定し、各ノード集合毎に、対応するエンティティの持つ属性を継承してその属性（所有者名、アクセス権等）を付与することはできない。登録した文書に関して分担執筆や部分構造の共有を行う場合、分担や共有の単位となるノード集合の設定と該ノード集合に対する属性設定が必要になるが、前記公知例 2 に示す従来技術を適用した文書管理システムではこれをユーザが手作業で行わなければならない、作業が極めて煩雑になる。

【0013】

また、前記公知例 2 に示す従来技術を適用した文書管理システムから登録済み文書を再びテキスト形式の SGML 文書として外部に取り出そうとした場合、登録時に存在していたエンティティ構成に対応する複数のファイルに分割して文書データを取り出すことはできず、文書内容全体を含む単一の文書エンティティとして取り出すことしかできない。従って、前記公知例 2 に示す従来技術を適用した文書管理システムを文書サーバとして使用し、該文書サーバにネットワークを介して接続したクライアントコンピュータで文書内容の表示や再利用を行うシステム構成をとった場合、クライアント側で必要とする部分的なテキストをエンティティ単位で取り出して処理することができず、毎回文書データ全体を取り出すことになるので、クライアントとサーバ間での転送データ量が増大し、処理効率が低下する。

【0014】

また、クライアント側で文書内容の編集・更新を行うシステム構成をとった場合、登録済み SGML 文書から特定のエンティティ部分だけを取り出して編集し、更新後のエンティティを再登録することにより文書論理構造を部分的に更新することができないので、複数人の担当者による大規模文書の共同管理・分担執筆が困難になる。

【0015】

更に、文書を構成するエンティティ毎に、その内容と作成者を保証する為に電子署名データを付加して管理しようとした場合、前記公知例2に示す従来技術を適用した文書管理システムでは文書を登録した段階でエンティティ単位のテキストデータは論理構造データに展開されて失われてしまう為、このような厳密な文書管理が不可能であるという問題がある。

【0016】

本発明の目的は上記問題を解決し、登録文書の持つ論理構造と実体構造とを相互に変換可能な形で対応付けて管理することが可能な技術を提供することにある。

【0017】

本発明の他の目的は属性設定を手作業で行うことなく登録前の構造化文書の各エンティティが持っていた属性に従って登録済み文書の各エンティティのアクセス制御を行うことが可能な技術を提供することにある。

【0018】

本発明の他の目的はエンティティ取り出し時の転送データ量を減少させて処理効率を向上させることが可能な技術を提供することにある。

【0019】

本発明の他の目的は複数人の担当者による大規模文書の共同管理・分担執筆を容易に行うことが可能な技術を提供することにある。

【0020】

本発明の他の目的はその内容と作成者を保証する為に文書を構成するエンティティ毎に電子署名データを付加して厳密な文書管理を行うことが可能な技術を提供することにある。

【0021】

【課題を解決するための手段】

本発明は、文書記述言語を用いて記述された構造化文書を管理する構造化文書管理システムにおいて、構造化文書の実体構成及び論理構造を示す対応関係データを使用して複数のエンティティから成る構造化文書を管理するものである。

【0022】

本発明では、SGML等の文書記述言語で記載された構造化文書を登録する際に登録対象文書の構文解析を行い、その実体構造を示すエンティティ構成表と、論理構造を示す解析済みインスタンスデータとを有する構文解析結果データを生成する。そして前記生成した構文解析結果データを参照し、登録対象文書を構成するエンティティのリソース名を当該エンティティのエンティティ識別子に対応付けて前記エンティティ構成表に設定する。すなわち登録対象文書を構成する全てのエンティティについて、そのエンティティ種別、エンティティ名、エンティティ識別子及びリソース名を表形式で前記エンティティ構成表にまとめ、登録対象文書の文書インスタンスを構文解析していく過程で検出したタグ等の記号で表されるイベントを、その出現順に従ってイベント種別とパラメタとの対応表の形で解析済みインスタンスデータにまとめておく。

【0023】

次に前記生成した構文解析結果データを参照して登録対象文書が持つ実体構造と論理構造との対応関係を抽出し、前記エンティティ識別子と解析済みインスタンスデータにより得られる論理構造部分とを対応付けて登録対象文書の実体構造及び論理構造を示す対応関係データを生成する。

【0024】

前記対応関係データは、登録対象文書の要素、内容データ及びエンティティ遷移をノードとして接続した木構造で表される。この対応関係データでは、あるエンティティ遷移ノードを根とする部分木から、下位のエンティティ遷移ノードを根とする全ての部分木を除いた部分が、そのエンティティ遷移ノードに対応するエンティティに内容として含まれる部分になる。すなわち本発明の対応関係データは、実体構造（エンティティ構成）と論理構造との対応関係をエンティティ遷移ノードという形でその内部に含んでいる。

【0025】

前記の様に本発明では登録対象文書が元々持っていた複数のエンティティの実体構造とそれらの論理構造を保持したまま登録を行っているので、登録した文書を更新する場合には登録済み文書の特定のエンティティまたは特定の論理構造部

分を選択した部分的な更新処理を行うことができる。

【0026】

登録済み文書のエンティティを更新する場合には登録済み文書の任意のエンティティを更新対象として選択し、前記更新対象の内容データを更新する。そして前記更新対象エンティティの実体構成及び論理構造を示す部分的な対応関係データを生成する。次に登録済み文書の対応関係データ中で前記更新対象エンティティに対応する部分範囲を求め、前記求めた部分範囲の対応関係データを前記更新対象エンティティの部分的な対応関係データによって置換して登録済み文書の対応関係データを更新する。

【0027】

登録済み文書の論理構造を更新する場合には登録済み文書の任意の論理構造を更新対象として選択し、前記更新対象の論理構造を編集する。そして前記編集の行われた論理構造の範囲内にあるエンティティの部分的な対応関係データを文字列化して記述した編集結果エンティティデータを生成し、前記生成した編集結果エンティティデータによって前記編集の行われた論理構造の範囲内のエンティティの内容データを置換する。次に登録済み文書の対応関係データ中で前記エンティティに対応する部分範囲を求め、前記求めた部分範囲の対応関係データを前記エンティティの部分的な対応関係データによって置換して登録済み文書の対応関係データを更新する。

【0028】

前記の様に本発明では、実体構造と論理構造とを対応付けて管理しており、木構造で表現された対応関係データと、文書記述言語を用いてテキスト形式で記述されたエンティティデータとの間で随時相互変換を行う。前記従来技術を適用した文書管理システムにSGML文書を登録した場合、複数のファイルから成るエンティティを文書内容全体が含まれた単一の文書エンティティとしてしまうので、登録前にそのSGML文書が持っていたエンティティ構成に関する情報は失われてしまうが、本発明では登録対象文書のエンティティ構成に関する情報が保持される。

【0029】

よって本発明で複数個のエンティティから構成されるSGML文書を登録して対応

関係データを生成する場合、登録対象文書の持つエンティティ構成に対応したノード集合を設定し、対応するエンティティの持つ属性を継承してその所有者名やアクセス権等の属性を各ノード集合毎に付与することができる。登録した文書に関して分担執筆や部分構造の共有を行う場合、分担や共有の単位となるノード集合の設定と該ノード集合に対する属性設定が必要になるが、本発明では対応するエンティティの持つ属性を継承するので、前記属性設定をユーザが手作業で行う必要がない。

【0030】

また登録済み文書を再びテキスト形式のSGML文書として外部に取り出そうとした場合、前記従来技術では登録時に存在していたエンティティ構成に対応する複数のファイルに分割して文書データを取り出すことはできないが、本発明では登録時に存在していたエンティティ構成をそのまま保持しているので、そのエンティティ構成に対応する複数のファイルに分割して文書データを取り出すことができる。

【0031】

従って本発明の文書管理システムを文書サーバとして使用し、該文書サーバにネットワークを介して接続したクライアントコンピュータで文書内容の表示や再利用を行うシステム構成をとった場合、クライアント側で必要とする部分的なテキストをエンティティ単位で取り出して処理することが可能で、毎回文書データ全体を取り出すことがないのでクライアントーサーバ間での転送データ量が減少して処理効率が向上する。

【0032】

またクライアント側で文書内容の編集・更新を行うシステム構成をとった場合、登録済みSGML文書から特定のエンティティ部分だけを取り出して編集し、更新後のエンティティを再登録することにより文書論理構造を部分的に更新することができるので、複数人の担当者による大規模文書の共同管理・分担執筆が容易に行える。

【0033】

更にエンティティ単位のテキストデータが失われてしまうことがないので、そ

の内容と作成者を保証する為に文書を構成するエンティティ毎に電子署名データを付加して厳密な文書管理を行うことが可能である。

【0034】

以上の様に本発明の構造化文書管理システムによれば、登録対象文書を構成する各エンティティと当該エンティティに対応する論理構造部分とを示す対応関係データを生成して複数のエンティティを管理するので、登録文書の持つ論理構造と実体構造とを相互に変換可能な形で対応付けて管理することが可能である。

【0035】

【発明の実施の形態】

以下に構造化文書の実体構造と論理構造とを対応付けて管理する一実施形態の構造化文書管理システムについて説明する。

図1は本実施形態の文書管理サーバの概略構成を示す図である。図1に示す様に本実施形態の文書管理サーバは、文書構文解析処理部421と、内容データ抽出処理部422と、対応関係データ抽出処理部423と、内容データ更新処理部424と、部分的対応関係データ生成処理部425と、対応部分範囲決定処理部426と、論理構造編集処理部427と、編集結果データ生成処理部428と、対応内容データ更新処理部429と、更新対象検出処理部430と、対応関係データ更新処理部431とを有している。

【0036】

文書構文解析処理部421は、登録対象文書の構文解析を行い、登録対象文書の実体構造を示すエンティティ構成表と、登録対象文書の論理構造を示す解析済みインスタンスデータとを有する構文解析結果データを生成する処理部である。

【0037】

内容データ抽出処理部422は、前記生成した構文解析結果データを参照し、登録対象文書を構成するエンティティの内容データを読み出して文書空間データベース格納領域418に格納し、前記格納したエンティティのリソース名を当該エンティティのエンティティ識別子に対応付けて前記エンティティ構成表に設定する処理部である。

【0038】

対応関係データ抽出処理部423は、登録対象文書を構成するエンティティと解析済みインスタンスデータにより得られる論理構造部分とを対応付けて登録対象文書の実体構造及び論理構造を示す対応関係データを生成する処理部である。

【0039】

内容データ更新処理部424は、登録済み文書の実体構造及び論理構造を示す対応関係データから登録済み文書を構成する任意のエンティティを更新対象として選択して更新対象エンティティの内容データを更新する処理部である。

【0040】

部分的対応関係データ生成処理部425は、前記更新対象エンティティの実体構造及び論理構造を示す部分的な対応関係データを生成する処理部である。対応部分範囲決定処理部426は、登録済み文書の対応関係データ中で前記更新対象エンティティに対応する部分範囲を求める処理部である。

【0041】

論理構造編集処理部427は、登録済み文書の実体構造及び論理構造を示す対応関係データから登録済み文書の任意の論理構造を更新対象として選択して編集する処理部である。編集結果データ生成処理部428は、前記編集した論理構造の範囲内にあるエンティティの実体構造及び論理構造を示す部分的な対応関係データを文字列化して記述した編集結果エンティティデータを生成する処理部である。

【0042】

対応内容データ更新処理部429は、前記生成した編集結果エンティティデータによって前記エンティティの内容データを置換する処理部である。更新対象検出処理部430は、登録済み文書の対応関係データ中で前記エンティティに対応する部分範囲を求める処理部である。対応関係データ更新処理部431は、前記求めた部分範囲の対応関係データを前記部分的な対応関係データによって置換して登録済み文書の対応関係データを更新する処理部である。

【0043】

文書管理サーバを文書構文解析処理部421、内容データ抽出処理部422、

対応関係データ抽出処理部 4 2 3、内容データ更新処理部 4 2 4、部分的対応関係データ生成処理部 4 2 5、対応部分範囲決定処理部 4 2 6、論理構造編集処理部 4 2 7、編集結果データ生成処理部 4 2 8、対応内容データ更新処理部 4 2 9、更新対象検出処理部 4 3 0 及び対応関係データ更新処理部 4 3 1 として機能させる為のプログラムは、CD-ROM等の記録媒体に記録され磁気ディスク等に格納された後、メモリにロードされて実行されるものとする。なお前記プログラムを記録する媒体はCD-ROM以外の他の媒体でも良い。

【0044】

なお文書登録制御プログラム 4 1 0 は、文書登録処理全体を制御すると共に、登録対象文書の持つエンティティ構成に対応したノード集合を設定し、対応するエンティティの持つ属性を継承してその属性を各ノード集合毎に付与したり、登録対象文書を構成するエンティティに付加された電子署名を当該エンティティの内容データと共に登録する処理を行う。

【0045】

また文書更新制御プログラム 4 1 2 は、登録済み文書の持つエンティティ構成に対応して設定されたノード集合の所有者名やアクセス権等の属性に応じてノード集合毎に更新対象へのアクセス制御を行い、登録済み文書を構成するエンティティ単位で編集対象を取り出して更新する処理や、複数の編集対象を別々に取り出して編集することにより構造化文書の編集を編集対象毎に並列に行う処理を制御し、更新後の編集対象を構成するエンティティに電子署名を付加する処理を行う。

【0046】

図 2 は本実施形態の文書管理システムの全体構成を示す図である。図 2 に示す通り本実施形態の文書管理システムは、文書管理サーバ 3 0 1、文書管理クライアント 3 0 2、文書業務処理システム 3 0 3、及びネットワーク 3 0 4 から構成される。

【0047】

文書管理サーバ 3 0 1 は、管理対象文書の登録・格納、登録済み文書の構造及び内容データに対する参照及び更新、特定の属性または内容文字列を持つ文書の

検索、登録済み文書の削除等、文書管理に関する主要な処理を実行する。

【0048】

文書管理クライアント302は、文書管理サーバ301が管理する文書群を操作する為のGUI(Graphical User Interface)を提供し、ユーザからの指示に従って文書の登録、参照、更新、検索等の処理を行う為の指令及びデータをネットワーク304を介して文書管理サーバ301に転送して前記指令を実行させる。

【0049】

業務処理システム303は、文書管理クライアント302と同様にユーザからの指示に従って指令及びデータを文書管理サーバ301に転送して該指令を実行させるが、一般的な文書管理用GUIではなく、特定の業務を実行する為の専用GUIをユーザに提供する。ネットワーク304は、局所ネットワーク(LAN:Local Area Network)または広域ネットワーク(WAN:Wide Area Network)であって、複数のコンピュータ装置を接続し、相互に指令及びデータを転送する手段を提供する。

【0050】

なお図2には文書管理クライアント302及び業務処理システム303の2台のコンピュータを文書管理サーバ301に対するクライアントとして使用する構成例を示したが、クライアントは1台のみの構成とすることも、3台以上とする構成をとることもできる。またクライアントは文書管理クライアントだけであっても良く、業務処理システムだけであっても良い。

【0051】

図2では文書管理クライアント302及び業務処理システム303から文書管理サーバ301に指令及びデータを転送する手段としてネットワーク304を使用するものとしたが、代わりにフロッピーディスク、光磁気ディスク、追記型光ディスク等の可搬型媒体を使用する構成をとることもでき、可搬型媒体とネットワーク304を併用する構成をとることもできる。或いはクライアントを文書管理サーバ301と同一のコンピュータ上で実行し、データ転送を行わない構成をとることもできる。

【0052】

以下、本実施形態の主要な構成要素である文書管理サーバ301について説明する。図1に示す文書管理サーバ301は、ディスプレイ401、キーボード402、中央処理装置403、フロッピーディスクドライブ404、フロッピーディスク405、通信制御装置406、主メモリ407、磁気ディスク装置408、システムバス409から構成される。

【0053】

ディスプレイ401は、本サーバにおける処理の実行状況等を表示する為に用いられる。キーボード402は、文書管理サーバ301の起動・終了等を指示するコマンドを入力する為に用いられる。中央処理装置403は、本サーバを構成する各種プログラムを実行する。フロッピーディスクドライブ404は、フロッピーディスク405に対するデータの読み書きの為に用いられる。

【0054】

フロッピーディスク405は、登録対象文書等、他のシステムとの間でデータ交換を行う為に用いられる。通信制御装置406は、ネットワーク304を介して文書管理クライアント302及び業務処理システム303と通信し、指令及びデータの交換を行う為に用いられる。

【0055】

主メモリ407は、本サーバにおける処理を実行する為の各種プログラム及び一時的なデータを保持する。磁気ディスク装置408は、登録された文書データ及び本サーバが文書管理用に使用する各種データを記憶する手段として用いられる。システムバス409は、これらの各種装置を相互に接続する。

【0056】

主メモリ407中には、文書登録制御プログラム410、文書参照制御プログラム411、文書更新制御プログラム412、文書出力制御プログラム413、文書検索制御プログラム414、文書空間管理プログラム415、データベース管理システム416及びオペレーティングシステム417が保持される。磁気ディスク装置408中には、文書空間データベース格納領域418、プログラム格納領域419及びデータファイル格納領域420が確保される。

【0057】

文書登録制御プログラム410は、文書記述言語SGMLを用いて記述された登録対象文書を、フロッピーディスク405、データファイル格納領域420、またはネットワーク304から読み込み、その構文解析を行い、登録対象文書の持つエンティティデータ、論理構造データ及びその両者の対応関係を文書空間データベース格納領域418に格納する。

【0058】

文書参照制御プログラム411は、文書空間データベース格納領域418に格納された登録済み文書のエンティティデータ、論理構造データまたはその両方を読み出し、それらのデータを構成する情報の一部を抽出して要求元クライアントに転送する。

【0059】

文書更新制御プログラム412は、文書空間データベース格納領域418に格納された登録済み文書のエンティティデータ、論理構造データまたはその両方に対して、そのデータ内容を更新する処理を行う。

【0060】

文書出力制御プログラム413は、文書空間データベース格納領域418から登録済み文書のエンティティデータを読み出し、該文書を構成するエンティティを、フロッピーディスク405またはデータファイル格納領域420に出力するか、或いはネットワーク304を介して要求元クライアント（文書管理クライアント302または業務処理システム303）に転送する。

【0061】

文書検索制御プログラム414は、要求元クライアントから受け取った検索条件式に従って文書空間データベース格納領域418に格納された登録済み文書群の検索を行い、検索結果データを要求元クライアントに転送する。

【0062】

文書空間管理プログラム415は、ネットワーク304を介して要求元クライアントからの指令を受信し、該指令の種別に対応して文書登録制御プログラム410、文書参照制御プログラム411、文書更新制御プログラム412、文書出

力制御プログラム413、または文書検索制御プログラム414を起動し、該指令の内容に従った処理を実行させる。また、登録済み文書群を分類整理する為のフォルダオブジェクトを生成して文書空間データベース格納領域418に格納する処理や、登録済み文書の一部を文書空間データベース格納領域418から削除する処理等、一部の処理については文書空間管理プログラム415自身が実行する。

【0063】

データベース管理システム416は、文書空間データベース格納領域418に対するデータの格納、更新、削除、検索に関する全ての処理を制御する。文書空間データベース格納領域418にアクセスする全てのプログラムは、その全ての処理をデータベース管理システム416を介して行う。オペレーティングシステム417は、システムの起動・初期化、各種周辺装置との間のデータ転送等、コンピュータ上で本サーバを構成する各プログラムを実行する為の基本機能を提供する。

【0064】

文書空間データベース格納領域418は、本サーバに登録された全ての文書に関するデータ、及び本サーバが文書管理を行う為に必要とする全ての関連データを格納するデータベースである。プログラム格納領域419は、本サーバを構成する全てのプログラムを格納しており、これらのプログラムはシステム起動時にオペレーティングシステム417によって読み出され、主メモリ407に転送される。データファイル格納領域420は、オペレーティングシステム417または本サーバ上で実行される各種ユーティリティプログラムが使用するデータファイルを格納する。

【0065】

なお本実施形態ではデータ交換の為の可搬型媒体としてフロッピーディスク405を使用する構成としたが、光磁気ディスクや追記型光ディスク等の他種の可搬型媒体を利用する構成をとることもできる。また本実施形態では文書データ及び文書管理用データをデータベース中に格納する構成としたが、データベース管理システムを使用せず、これらのデータをファイルとしてデータファイル格納領

域420に格納する構成をとることもできる。

【0066】

次に本実施形態における文書管理サーバ301の基本動作手順についてPAD (Problem Analysis Diagram) を用いて説明する。

図3は本実施形態の文書管理サーバ301全体の動作を制御する文書空間管理プログラム415の概略処理手順を示すPAD図である。キーボード402からのサーバ起動コマンド等により文書空間管理プログラム415が起動されると、本プログラムはサーバとしてクライアントから指令を受信してはその処理を行うループに入る(ステップ501)。このループは、キーボード402からサーバの停止を指示するコマンドが入力されるまで継続する。

【0067】

ステップ501のループは、クライアントからの指令を受信する処理(ステップ502)と、受信した指令の種別を判定し、該種別に対応する処理に分岐する処理(ステップ503)を繰り返す。

【0068】

ステップ503では、受信した指令の種別を判定し、該指令の種別に応じてステップ504からステップ511までのいずれかの処理を選択して実行する。ここで、前記指令が新たな文書の登録を求める指令であった場合、文書登録制御プログラム410を実行して指定された文書を文書空間データベース格納領域418に格納する(ステップ504)。

【0069】

また前記指令が特定の登録済み文書の持つ内容データへの参照を求める指令であった場合、文書参照制御プログラム411を実行して要求されたデータを文書空間データベース格納領域418から読み出し、該データを要求元クライアントに返信する(ステップ505)。

【0070】

また前記指令が特定の登録済み文書の持つ内容データの更新を求める指令であった場合、文書更新制御プログラム412を実行して文書空間データベース格納領域418に格納された登録済み文書の内容データを更新する(ステップ506)

）。

【0071】

また前記指令が特定の登録済み文書の一部または全体を外部に出力することを求める指令であった場合、文書出力制御プログラム413を実行して文書空間データベース格納領域418から登録済み文書の内容データを読み出し、前記指令中の指示に従って、該データをフロッピーディスク405またはデータファイル格納領域420に出力するか或いはネットワーク304を介して要求元クライアントに転送する（ステップ507）。

【0072】

また前記指令が特定の検索条件式を満たす文書の集合を求める指令であった場合、文書検索制御プログラム414を実行して文書空間データベース格納領域418に格納された登録済み文書群の検索を行い、検索結果データを要求元クライアントに転送する（ステップ508）。

【0073】

また前記指令が登録済み文書群を分類整理する為のフォルダオブジェクトの生成を求める指令であった場合、新たなフォルダオブジェクトを生成して文書空間データベース格納領域418に格納する（ステップ509）。

【0074】

また前記指令が特定の文書の特定のフォルダオブジェクトへの関連付け、または特定の関連付けの削除を求める指令であった場合、要求された関連付けの追加または削除を行う（ステップ510）。

【0075】

また前記指令が特定の登録済み文書または特定のフォルダオブジェクトを削除することを求める指令だった場合、文書空間データベース格納領域418から指定された文書またはフォルダオブジェクトを削除する（ステップ511）。

【0076】

図4は本実施形態の文書登録制御プログラム410の概略処理手順を示すPAD図である。本プログラムは、文書の登録を求めるクライアントからの指令に対応して文書空間管理プログラム415によって呼び出される。以下に本実施形態に

おける文書登録制御プログラム410の動作手順を説明する。

【0077】

本プログラムは、文書空間管理プログラム415によって呼び出されると、まず登録対象文書を構成するエンティティの一覧を呼び出してパラメタとして入手する（ステップ601）。ここで該一覧には、登録対象文書を構成する各エンティティの内容データを保持しているリソースのリソース名と、登録対象文書が複数のエンティティから構成される場合にそのいずれが文書エンティティであるかを識別する情報を含んでいる。

【0078】

エンティティの内容データの入手元となるリソースとしては、フロッピーディスク405またはデータファイル格納領域420上に格納されたデータファイル、ネットワーク304を介してアクセスできるネットワークリソース及び文書空間データベース格納領域418中に格納されている登録済みエンティティが使用できる。リソース名としては、リソースがデータファイルである場合にはそのパス名、ネットワーク上のリソースである場合にはそのネットワークアドレス、登録済みエンティティである場合には該エンティティを他の登録済みエンティティから区別して一意に識別するエンティティ識別子を指定する。

【0079】

次に登録対象文書に対して該文書を全ての登録済み文書から区別して一意に識別する文書識別子を割り当て（ステップ602）、文書構文解析処理部421により登録対象文書の文書エンティティから出発してその構文解析を行い、構文解析結果データを生成する（ステップ603）。なおこの構文解析の過程で、登録対象文書を構成するエンティティの内、登録済みエンティティを除く全てのエンティティに対して、該エンティティを全ての登録済みエンティティから区別して一意に識別するエンティティ識別子を割り当てる。

【0080】

次に内容データ抽出処理部422は、前記生成した構文解析結果データを参照して、登録対象文書を構成するエンティティの内、登録済みエンティティを除く全てのエンティティについて、ステップ605及びステップ606を繰り返し実

行する（ステップ604）。ここでステップ605では処理対象となるエンティティの内容データを対応するリソースから読み出す。ステップ606では処理対象となるエンティティのエンティティ識別子とその内容データとの組を文書空間データベース格納領域418に格納してエンティティの登録を行う。

【0081】

次にステップ607で対応関係データ抽出処理部423は、前記構文解析結果データを参照し、登録対象文書が持つ実体構造と論理構造との対応関係を抽出し、登録対象文書を構成する各エンティティと当該エンティティに対応する論理構造部分とを示す対応関係データを生成する。ステップ608では、登録対象文書に割り当てた文書識別子と前記対応関係データとの組を文書空間データベース格納領域418に格納して文書の登録を行い、処理を終了する。

【0082】

図5は本実施形態の構文解析結果データのデータ構造を示す図である。図5において文書登録制御プログラム410が前記ステップ603で生成する構文解析結果データ701は、二つの主要な部分、すなわちエンティティ構成表702と解析済みインスタンスデータ703とからなる。

【0083】

エンティティ構成表702は、登録対象文書を構成する全てのエンティティについて、そのエンティティ種別、エンティティ名、エンティティ識別子及びリソース名を表形式でまとめたデータである。ここで項目「エンティティ種別」は、そのエンティティが文書エンティティ(DOC)、テキストエンティティ(TEXT)または非SGMLデータエンティティ(NDATA)のいずれであるかを記号によって示している。項目「エンティティ名」は、登録対象文書中でそのエンティティを参照する為に用いられているエンティティ名を文字列として示している。但し、そのエンティティが文書エンティティであった場合、本項目は空欄となる。項目「エンティティ識別子」は、そのエンティティに割り当てられたエンティティ識別子を示している。項目「リソース名」は、そのエンティティの内容データを保持するリソースのリソース名を文字列として示している。但し、そのエンティティが登録済みエンティティであった場合、本項目は空欄となる。

【0084】

解析済みインスタンスデータ703は、登録対象文書の文書インスタンスを構文解析していく過程で検出したイベントを、その出現順に従ってイベント種別とパラメタとの対応表の形でまとめたデータである。ここで、項目「イベント種別」は、そのイベントが要素の開始タグ(STAG)、終了タグ(ETAG)、内容を持たない要素の開始と終了を兼ねるタグ(ITAG)、解析済み文字列データ(PCDATA)、エンティティ参照(EREF)、エンティティ終了(EE)のいずれであるかを記号によって示している。項目「パラメタ」は、イベント種別に応じて決まる付加的なデータを示している。イベント種別が開始タグである場合、その開始タグを先頭位置とする要素の要素型名、及び該開始タグ内で指定された属性値（オプション）がパラメタとして示される。イベント種別が終了タグである場合、その終了タグを末尾位置とする要素の要素型名がパラメタとして示される。イベント種別が解析済み文字列データである場合、該文字列の内容がパラメタとして示される。イベント種別がエンティティ参照である場合、被参照エンティティのエンティティ名がパラメタとして示される。エンティティ終了イベントはパラメタを持たない。

【0085】

図6は本実施形態のエンティティ構成表702の具体例を示す図である。図6では図15に示したSGML文書を構文解析した結果として生成されるエンティティ構成表702の例を表している。図6における行801～804が、それぞれ図15に示した文書エンティティ101～データエンティティ104に対応する。

【0086】

図7は本実施形態の解析済みインスタンスデータ703の具体例を示す図である。図7では図15に示したSGML文書を構文解析した結果として生成される解析済みインスタンスデータ703の例を表している。図7に示す通り、文書記述言語SGMLの構文に従って記述されたSGML文書は、前記ステップ603により、タグ、解析済み文字列データ、エンティティ参照等の出現を示すイベントの列に変換される。

【0087】

文書登録制御プログラム410の対応関係データ抽出処理部423は、前記構

文解析結果データの解析済みインスタンスデータ703を参照して登録対象文書が持つ論理構造を抽出すると共に、解析済みインスタンスデータ703中のエンティティ参照(EREF)とエンティティ構成表702中のエンティティ識別子とを対応付け、登録対象文書を構成する各エンティティのエンティティ識別子と当該エンティティに対応する論理構造部分とを木構造で表した対応関係データを生成する。

【0088】

図8は本実施形態の対応関係データのデータ構造を示す図である。図8では文書登録制御プログラム410の対応関係データ抽出処理部423が生成する対応関係データのデータ構造を表している。この図では、図15に示すSGML文書を登録対象文書として処理した場合に生成される対応関係データを例として示している。

【0089】

図8に示す通り、対応関係データ1001は木構造として表現される。図8において楕円形のノードは要素を表しており、ノード中の文字列はその要素の要素型を示している。また矩形のノードは内容データを表しており、ノード中の文字列は内容データの種別（解析済み文字列データPCDATA、または非SGMLデータNDATA）を示している。菱形のノードはエンティティ遷移を表しており、ノード中の文字列はエンティティ識別子を示している。

【0090】

ここで、あるエンティティ遷移ノードを根とする部分木から、下位のエンティティ遷移ノードを根とする全ての部分木を除いた部分が、そのエンティティ遷移ノードに対応するエンティティに内容として含まれる部分になる。すなわち本実施形態の文書登録制御プログラム410が生成する対応関係データは、論理構造と実体構造（エンティティ構成）との対応関係を、エンティティ遷移ノードという形でその内部に含んでいる。図8の例では、鎖線で囲った部分範囲1002が、エンティティE0002（図6に示すテキストエンティティ802）の内容に対応することになる。

【0091】

図9は本実施形態の文書更新制御プログラム412の概略処理手順を示すPAD図である。本プログラムは、文書の更新を求めるクライアントからの指令に対応して、文書空間管理プログラム415によって呼び出される。以下に本実施形態における文書更新制御プログラム412の動作手順を説明する。

本プログラムは、文書空間管理プログラム415によって呼び出されると、まず更新対象として登録済み文書を構成するいずれかのエンティティの内容が指定されているのか、或いは登録済み文書の論理構造が指定されているのかを判定する（ステップ1101）。ここで更新対象としてエンティティが指定されている場合、ステップ1102及びステップ1103からなる処理に分岐する。また更新対象として論理構造が指定されている場合、ステップ1104及びステップ1105からなる処理に分岐し、分岐先から復帰した時点で処理を終了する。

ステップ1102では、呼び出し時に与えられたパラメタ（エンティティ識別子）を用いて、更新対象となる登録済みエンティティを選択する。次にステップ1103において、前記選択されたエンティティの内容データを更新する処理を行う。

ステップ1104では、呼び出し時に与えられたパラメタ（文書識別子）を用いて、更新対象となる論理構造を保持する登録済み文書を選択する。次にステップ1105において、前記選択された文書の持つ論理構造を更新する処理を行う。

【0092】

図10は本実施形態の選択されたエンティティの内容を更新する処理の手順を示すPAD図である。以下に図9におけるステップ1103、すなわち選択されたエンティティの内容を更新する処理について、その詳細を説明する。

【0093】

図10に示す様に被選択エンティティの更新を行う内容データ更新処理部424では、複数のユーザによって1個のエンティティが同時に更新されることを防ぐ為、まず更新対象エンティティに対してロックを設定する（ステップ1201）。次にロックの設定に成功したかどうかをチェックし（ステップ1202）、

成功した場合にはステップ1203からステップ1211までの更新処理を実行する。失敗した場合には、既にロックが設定されていて更新ができない旨のエラーコードを返却して処理を終了する（ステップ1213）。

【0094】

ステップ1203では、更新対象エンティティの内容データを文書空間データベース格納領域418から読み出す。次にステップ1204では、読み出した内容データを、外部プログラムを用いて編集可能な外部領域（フロッピーディスク405、データファイル格納領域420、またはネットワーク304を介して接続されたクライアント側のワーク領域等）に出力する。そして一旦処理を中断し、出力した内容データの外部プログラム（例えばSGMLエディタ）による編集の終了を待つ。外部プログラムによる編集処理が終了した旨の通知を受けると処理を再開し、ステップ1205に進む。ステップ1205では、更新後の内容データを外部領域から読み込み、該読み込んだ内容データによって更新対象エンティティの内容データを置換して文書空間データベース格納領域418に格納する。

【0095】

次にステップ1206で部分的対応関係データ生成処理部425は、更新対象エンティティの種別を判定する。更新対象エンティティが文書エンティティまたはテキストエンティティである場合にはステップ1207及びステップ1208を実行し、それ以外の非SGMLデータエンティティである場合にはステップ1209を実行する。

【0096】

ここでステップ1207では、更新対象エンティティの更新された内容データの構文解析を行い、前記更新対象エンティティのエンティティ構成及び解析済みインスタンスデータを示す構文解析結果データを生成する。ステップ1208では、該構文解析結果データを参照して前記更新対象エンティティを構成する各エンティティと解析済みインスタンスデータにより得られる論理構造部分とを対応付けて部分的な対応関係データを生成する。またステップ1209では、更新された内容データそのものを前記の部分的な対応関係データとして設定する。

【0097】

次にステップ1210で対応部分範囲決定処理部426は、全ての登録済み文書の対応関係データに対する検索を行い、更新対象エンティティに対応する（更新対象エンティティのエンティティ識別子を保持する）全てのエンティティ遷移ノードを検出する。そして検出された全てのエンティティ遷移ノードに関して、ステップ1211を繰り返し実行する。ここでステップ1211で対応関係データ更新処理部431は、選択されたエンティティ遷移ノードの下位にある部分的な対応関係データをステップ1208またはステップ1209で設定された部分的な対応関係データによって置換し、これを文書空間データベース格納領域中に格納することによって、更新されたエンティティに対応する登録済み文書の対応関係データを更新する。最後にステップ1212において更新対象エンティティに設定したロックを解除して処理を終了する。

【0098】

図11は本実施形態の選択されたエンティティの内容を更新する処理の経過例を示す図である。図11では図10にその手順を示した選択されたエンティティの内容を更新する処理についてその具体的な経過の例を表しており、図11では図15に示したテキストエンティティ103の内容を外部領域に出力し、外部プログラムによって編集する場合について示している。

【0099】

ここではテキストエンティティ103中の要素 note に含まれる内容文字列の一部をタグ及びで囲み、独立した要素 em に変更したものとする。この編集処理によってテキストエンティティ103の内容は編集結果エンティティデータ1301の様に更新される。このエンティティはテキストエンティティであるから、前記ステップ1207において構文解析が行われ、ステップ1208において部分的な対応関係データが生成される。この例の場合、生成される対応関係データは対応関係データ1302に示す構造を持つことになる。

【0100】

次にステップ1210において、更新対象エンティティに対応するエンティティ遷移ノードが検索される。図11に示す例の場合、登録された文書は図8に示

した対応関係データ1001を持つので、この木構造上では前記更新されたテキストエンティティ103（エンティティ識別子E0003）に対応するエンティティ遷移ノード2個（対応関係データ1001中の着色部分の上端）が検出される。従ってステップ1211において、これらのエンティティ遷移ノードを根とする部分構造の置き換えが行われ、この文書の対応関係データは最終的に対応関係データ1303に示す様に更新される。

【0101】

図12は本実施形態の選択された文書の論理構造を更新する処理の手順を示すPAD図である。以下に図9におけるステップ1105、すなわち選択された文書の対応関係データを更新する処理について、その詳細を説明する。

【0102】

図12に示す様に被選択文書の論理構造の更新を行う論理構造編集処理部427では、まず編集範囲の選択を行う（ステップ1401）。編集範囲の選択は、編集しようとする文書の対応関係データの論理構造上で、任意のノードを指定することによって行う。指定されたノードを根とする部分木が編集可能な範囲となる。

【0103】

次にステップ1402からステップ1406までの処理では、選択された編集範囲に対応する全てのエンティティについて、複数ユーザによる重複更新を防ぐ為にロックを設定する。ここで選択範囲の内部に含まれるエンティティ遷移ノード、及び選択範囲より上位でかつ最も近い位置にあるエンティティ遷移ノードに対応するエンティティが、その選択範囲に対応するエンティティとなる。

【0104】

ステップ1403では、現在注目している特定のエンティティについて更新ロックを設定する。次にロックの設定に成功したかどうかをチェックし（ステップ1404）、失敗した場合には設定済みのロックを全て解除した上で（ステップ1405）、既にロックが設定されていて更新ができない旨のエラーコードを返却して処理を終了する（ステップ1406）。

【0105】

対応エンティティ全てへのロックの設定が成功した場合にはステップ1407に進む。ステップ1407ではGUI等を介したユーザからの指示に従い、対応関係データの論理構造に対する編集操作（下位ノードの追加、削除、文字列データの変更等）を行う。このとき、変更を加えた位置を含むエンティティには内容を変更したことを示すマークを付加しておく。

【0106】

編集操作の終了が指示されるとステップ1408に進む。ステップ1408では、編集操作の過程で更新マークが付加された全てのエンティティについてステップ1409からステップ1414までの処理を繰り返す。

【0107】

ステップ1409で編集結果データ生成処理部428は、現在注目しているエンティティに対応する範囲の部分的な対応関係データを入手し、これを一時記憶に格納する。ステップ1410では、ステップ1409で入手した部分的な対応関係データをSGMLの構文に従って文字列化して記述した編集結果エンティティデータの生成を行う。次にステップ1411で対応内容データ更新処理部429は、この編集結果エンティティデータによって注目エンティティの内容データを置換して文書空間データベース格納領域418に格納する。

【0108】

ステップ1412で更新対象検出処理部430は、全ての登録済み文書の対応関係データに対する検索を行い、更新された注目エンティティに対応する（注目エンティティのエンティティ識別子を保持する）全てのエンティティ遷移ノードを検出する。そして検出された全てのエンティティ遷移ノードに関してステップ1413及びステップ1414の処理を繰り返し実行する。

【0109】

ステップ1413で対応関係データ更新処理部431は、注目するエンティティ遷移ノードが、ステップ1407で論理構造の編集を行った範囲に対応するものであるかどうかを判定し、編集済みのものでない場合には、ステップ1409で入手した更新後の部分的な対応関係データのコピーを生成して、これによって

注目するエンティティ遷移ノードを根とする部分木を置換する（ステップ1414）。

【0110】

ステップ1408における繰り返しが終了するとステップ1415に進み、変更が加えられた全ての文書について、その対応関係データを文書空間データベース格納領域418に格納する。最後に、エンティティ群に設定していたロックを全て解除し（ステップ1416）、処理を終了する。

【0111】

図13は本実施形態の選択された文書の対応関係データを更新する処理の経過例を示す図である。図13では図12にその手順を示した選択された文書の対応関係データを更新する処理についてその具体的な経過の例を表しており、図15に示したSGML文書の対応関係データを編集する場合を示している。

【0112】

対応関係データ1001は、編集を行う前の論理構造であり、図8における対応関係データ1001と同一である。ここでsec型要素1501を指定して編集範囲の選択を行ったとすると、本図に示す様にsec型要素1501を根とする部分木が編集可能範囲となる。その結果、編集範囲に含まれるエンティティ遷移ノードに対応するエンティティE0003、及び編集範囲の上位に位置し、かつ最も近い位置にあるエンティティ遷移ノードに対応するエンティティE0002では、編集範囲に対応するエンティティとしてロックが設定される。

【0113】

ロックの設定後編集操作が行われるが、ここではnote型要素1502の下位にある解析済み文字列データを分割し、その一部をem型要素とする構造編集が行われたものとする。この場合、更新後における文書全体の論理構造は、対応関係データ1503のようになる。編集操作が最終的にこれだけだったとすると、内容を変更したとしてマークを付加されるエンティティは、E0003だけとなる。そして、更新後の部分的な対応関係データから、該データをSGMLの構文に従って文字列化した編集結果エンティティデータが生成される。本図に示す例では、図11に示す編集結果エンティティデータ1301と同一の文字列が編集結果エンティティ

ィデータとして得られたものとする。

【0114】

エンティティE0003の内容データが編集結果エンティティデータ1301によって置換され、該エンティティが更新された後、該エンティティに対応する全てのエンティティ遷移ノード（但し、構造編集を行った範囲に対応するものを除く）の下位構造が、構造編集後の部分的な対応関係データのコピーによって置換される。本図の例では、エンティティE0003に対応するエンティティ遷移ノードが文書中にもう一箇所含まれている為、この部分の構造の置換が行われ、最終的な更新後の対応関係データは図11に示した対応関係データ1303と同一になる。

【0115】

以上に示した様に本実施形態における文書の更新処理では、文書を構成するエンティティをテキストとして編集した場合でも、文書の論理構造を直接構造編集した場合でも、実質的な編集内容が等価であれば最終的に論理構造とエンティティの内容データの間で整合性のとられた同一の結果が残される。なお本実施形態ではSGMLを用いて記述された構造化文書を登録対象文書として処理する構成をとったが、XMLやHTML等、他の文書記述言語を用いて記述された構造化文書を管理対象とする様に構成することもできる。

【0116】

図14は本実施形態の文書登録制御プログラム410が生成する対応関係データの他のデータ構造を示す図である。この図では図15に示すSGML文書を登録対象文書として処理した場合に生成される対応関係データを例として示している。以下に本実施形態の対応関係データの他の例について図面を用いて説明する。

【0117】

図14に示す様に対応関係データ1601は、図8に示した対応関係データ1001とは異なり、木構造中にエンティティ遷移ノードを含んでいない。独立したノードとしてエンティティ遷移ノードを保持する代わりに、対応関係データ1601では論理構造を構成する個々の要素ノード及び内容データノードが、それぞれ対応するエンティティのエンティティ識別子をプロパティとして保持してい

る。

【0118】

全ての要素ノード及び内容データノードにエンティティ識別子を持たせる為、対応関係データ1601のデータ量は、前記の対応関係データ1001の場合に比べて大きくなる。その代わり、図12に示した選択された文書の論理構造を更新する処理において編集範囲に対応するエンティティを求める際に、ノードの木構造を上方に辿ってエンティティ遷移ノードを検索する必要がなく、編集範囲に含まれるノードがプロパティとして持つエンティティ識別子の集合を得るだけで良いので処理が簡略化され、処理効率も向上する。なお対応関係データ1601を用いた場合と対応関係データ1001を用いた場合との間で機能的な差異は生じない。

【0119】

以上説明した様に本実施形態の構造化文書管理システムによれば、登録対象文書を構成する各エンティティと当該エンティティに対応する論理構造部分とを示す対応関係データを生成して複数のエンティティを管理するので、登録文書の持つ論理構造と実体構造とを相互に変換可能な形で対応付けて管理することが可能である。

【0120】

また本実施形態の構造化文書管理システムによれば、登録済み文書を構成するエンティティが元々持っていた属性を継承するので、属性設定を手作業で行うことなく登録前の構造化文書の各エンティティが持っていた属性に従って登録済み文書の各エンティティのアクセス制御を行うことが可能である。

【0121】

また本実施形態の構造化文書管理システムによれば、登録済み文書をエンティティ単位で部分的に取り出して処理するので、エンティティ取り出し時の転送データ量を減少させて処理効率を向上させることが可能である。

【0122】

また本実施形態の構造化文書管理システムによれば、登録済み文書から複数の編集対象をエンティティ単位で別々に取り出して編集を行うので、複数人の担当

者による大規模文書の共同管理・分担執筆を容易に行うことが可能である。

【0123】

また本実施形態の構造化文書管理システムによれば、エンティティ単位のテキストデータが失われてしまうことがないので、その内容と作成者を保証する為に文書を構成するエンティティ毎に電子署名データを付加して厳密な文書管理を行うことが可能である。

【0124】

【発明の効果】

本発明によれば登録対象文書を構成する各エンティティと当該エンティティに対応する論理構造部分とを示す対応関係データを生成して複数のエンティティを管理するので、登録文書の持つ論理構造と実体構造とを相互に変換可能な形で対応付けて管理することが可能である。

【図面の簡単な説明】

【図1】

本実施形態の文書管理サーバの概略構成を示す図である。

【図2】

本実施形態の文書管理システムの全体構成を示す図である。

【図3】

本実施形態の文書管理サーバ301全体の動作を制御する文書空間管理プログラム415の概略処理手順を示すPAD図である。

【図4】

本実施形態の文書登録制御プログラム410の概略処理手順を示すPAD図である。

【図5】

本実施形態の構文解析結果データのデータ構造を示す図である。

【図6】

本実施形態のエンティティ構成表702の具体例を示す図である。

【図7】

本実施形態の解析済みインスタンスデータ703の具体例を示す図である。

【図8】

本実施形態の対応関係データのデータ構造を示す図である。

【図9】

本実施形態の文書更新制御プログラム412の概略処理手順を示すPAD図である。

【図10】

本実施形態の選択されたエンティティの内容を更新する処理の手順を示すPAD図である。

【図11】

本実施形態の選択されたエンティティの内容を更新する処理の経過例を示す図である。

【図12】

本実施形態の選択された文書の論理構造を更新する処理の手順を示すPAD図である。

【図13】

本実施形態の選択された文書の対応関係データを更新する処理の経過例を示す図である。

【図14】

本実施形態の文書登録制御プログラム410が生成する対応関係データの他のデータ構造を示す図である。

【図15】

従来の文書記述言語SGMLによって記述された構造化文書の一例を示す図である。

【図16】

従来の図15のSGML文書を解析して得られる論理構造の例を示す図である。

【符号の説明】

401…ディスプレイ、402…キーボード、403…中央処理装置、404…フロッピーディスクドライブ、405…フロッピーディスク、406…通信制御装置、407…主メモリ、408…磁気ディスク装置、409…システムバス

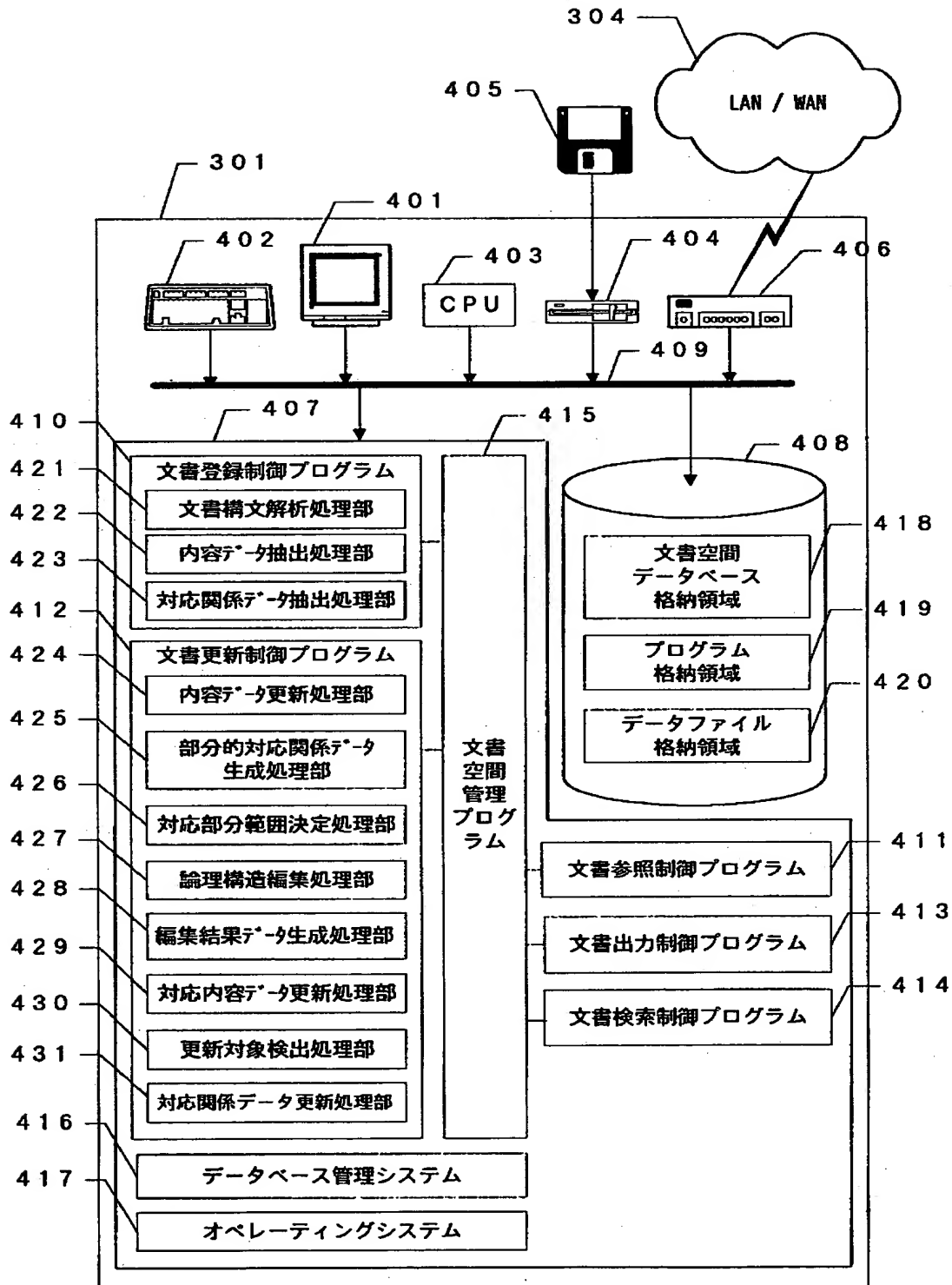
、410…文書登録制御プログラム、411…文書参照制御プログラム、412…文書更新制御プログラム、413…文書出力制御プログラム、414…文書検索制御プログラム、415…文書空間管理プログラム、416…データベース管理システム、417…オペレーティングシステム、418…文書空間データベース格納領域、419…プログラム格納領域、420…データファイル格納領域、421…文書構文解析処理部、422…内容データ抽出処理部、423…対応関係データ抽出処理部、424…内容データ更新処理部、425…部分的対応関係データ生成処理部、426…対応部分範囲決定処理部、427…論理構造編集処理部、428…編集結果データ生成処理部、429…対応内容データ更新処理部、430…更新対象検出処理部、431…対応関係データ更新処理部、301…文書管理サーバ、302…文書管理クライアント、303…業務処理システム、304…ネットワーク、701…構文解析結果データ、702…エンティティ構成表、703…解析済みインスタンスデータ、801～804…行、1001…対応関係データ、1002…部分範囲、1301…編集結果エンティティデータ、1302及び1303…対応関係データ、1501…sec型要素、1502…note型要素、1503…対応関係データ、1601…対応関係データ、101…文書エンティティ、102及び103…テキストエンティティ、104…データエンティティ、201…ノード集合、202…ノード集合、203及び204…ノード集合、205…ノード集合。

【書類名】

図面

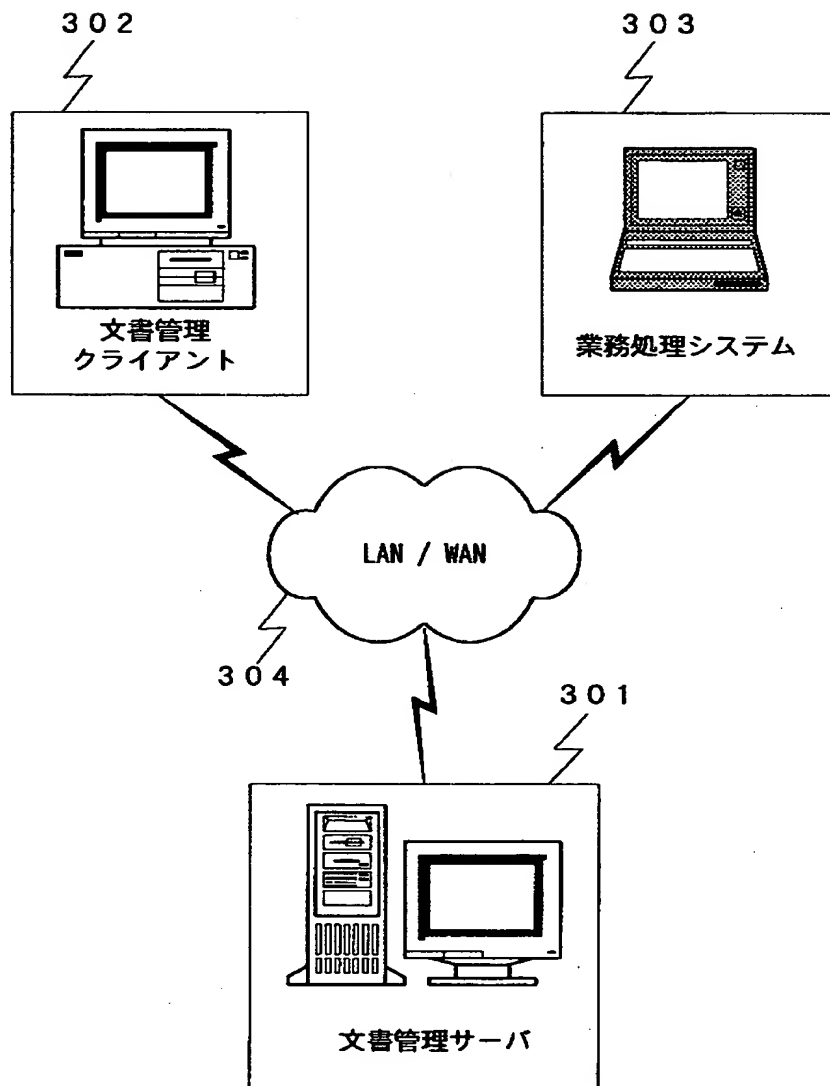
【図1】

図1



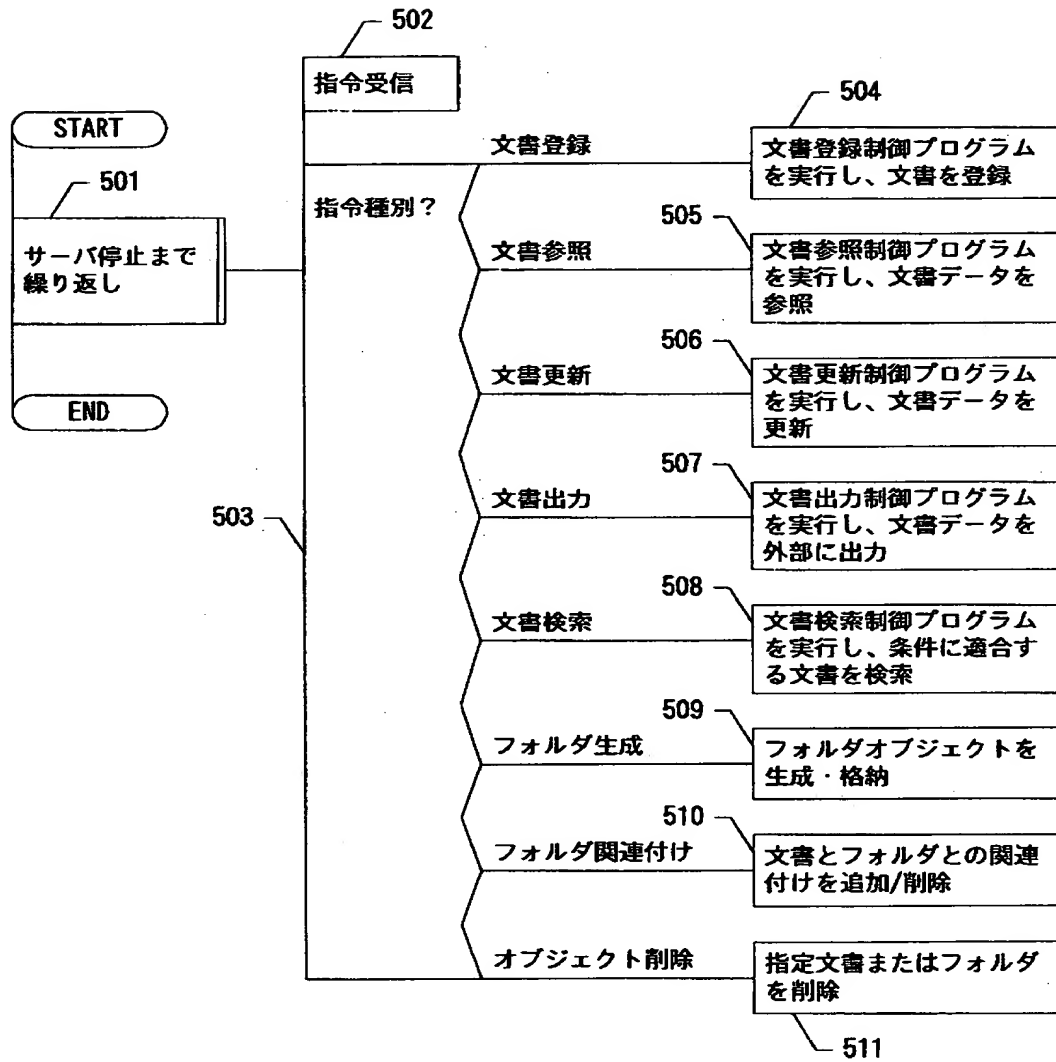
【図 2】

図 2



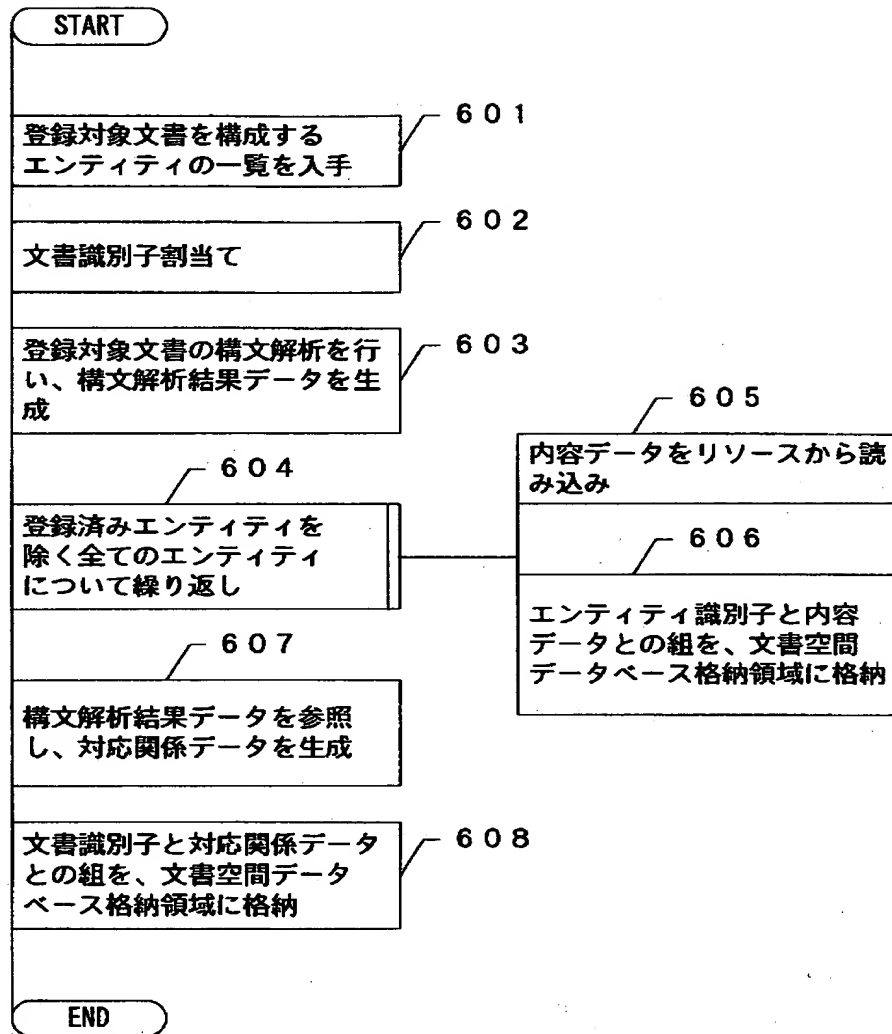
【図 3】

図 3



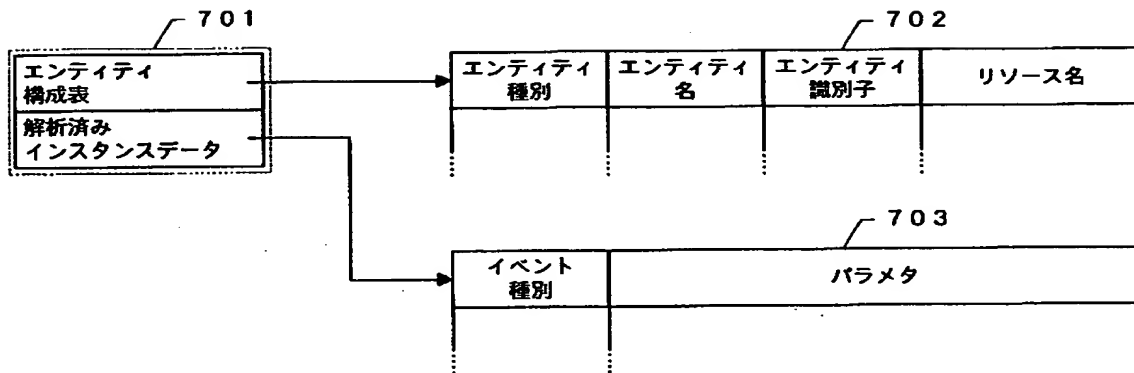
【図4】

図4



【図 5】

図 5



【図 6】

図 6

| エンティティ 種別 | エンティティ 名 | エンティティ 識別子 | リソース名 | |
|--------------|-------------|---------------|-----------|-----|
| DOC | -- | E0001 | doc.sgm | 801 |
| TEXT | part1 | E0002 | part1.sgm | 802 |
| TEXT | note | E0003 | note.sgm | 803 |
| NDATA | Fig1 | E0004 | fig1.bmp | 804 |

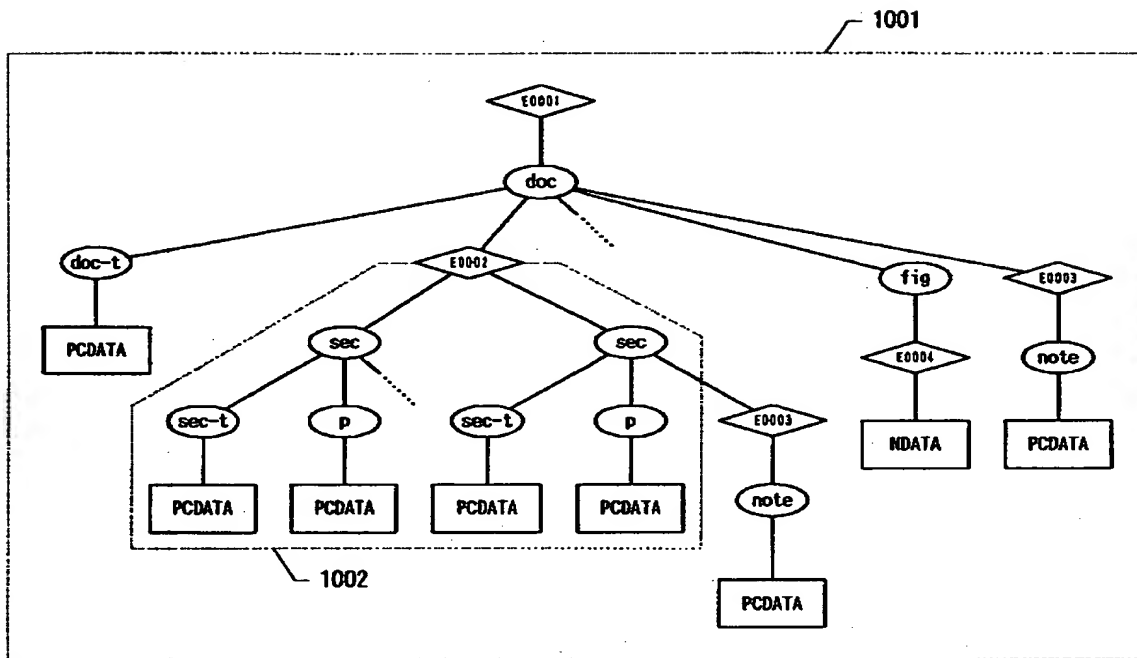
【図7】

図7

| イベント 種別 | パラメタ |
|------------|-------------------------------|
| STAG | type="doc" |
| STAG | type="doc-t" |
| PCDATA | text="文書検索技術" |
| ETAG | type="doc-t" |
| EREF | ent="part1" |
| STAG | type="sec" |
| STAG | type="sec-t" |
| PCDATA | text="キーワード検索" |
| ETAG | type="sec-t" |
| STAG | type="p" |
| PCDATA | text="あらかじめ用意したキーワードに適合…" |
| ETAG | type="p" |
| ETAG | type="sec" |
| STAG | type="sec" |
| STAG | type="sec-t" |
| PCDATA | text="全文検索" |
| ETAG | type="sec-t" |
| STAG | type="p" |
| PCDATA | text="ここでいう全文検索とは、登録された…" |
| ETAG | type="p" |
| EREF | ent="note" |
| STAG | type="note" |
| PCDATA | text="全文検索技術一般については、[文献1]…" |
| ETAG | type="note" |
| EE | |
| ETAG | type="sec" |
| EE | |
| ITAG | type="fig" attr-entity="Fig1" |
| EREF | ent="note" |
| STAG | type="note" |
| PCDATA | text="全文検索技術一般については、[文献1]…" |
| ETAG | type="note" |
| EE | |
| ETAG | type="doc" |

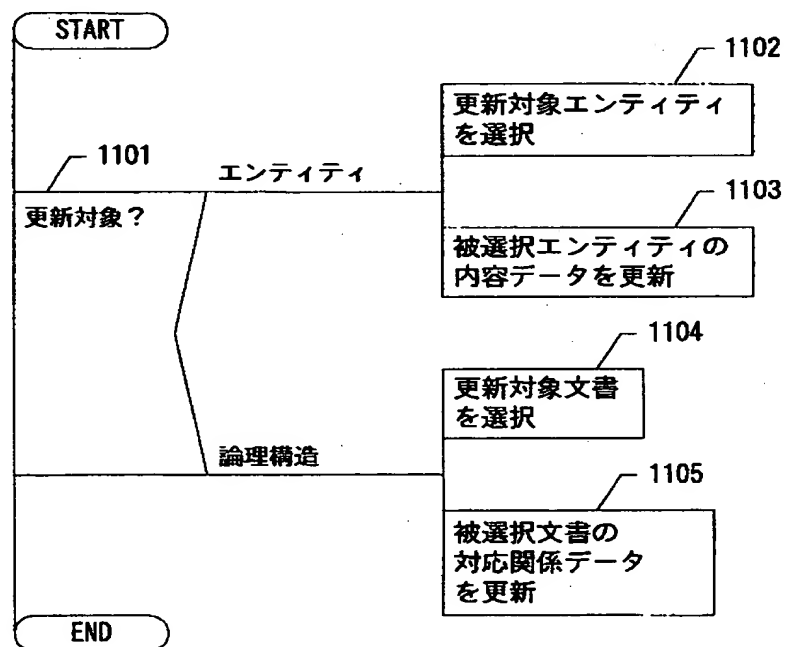
【図 8】

図 8



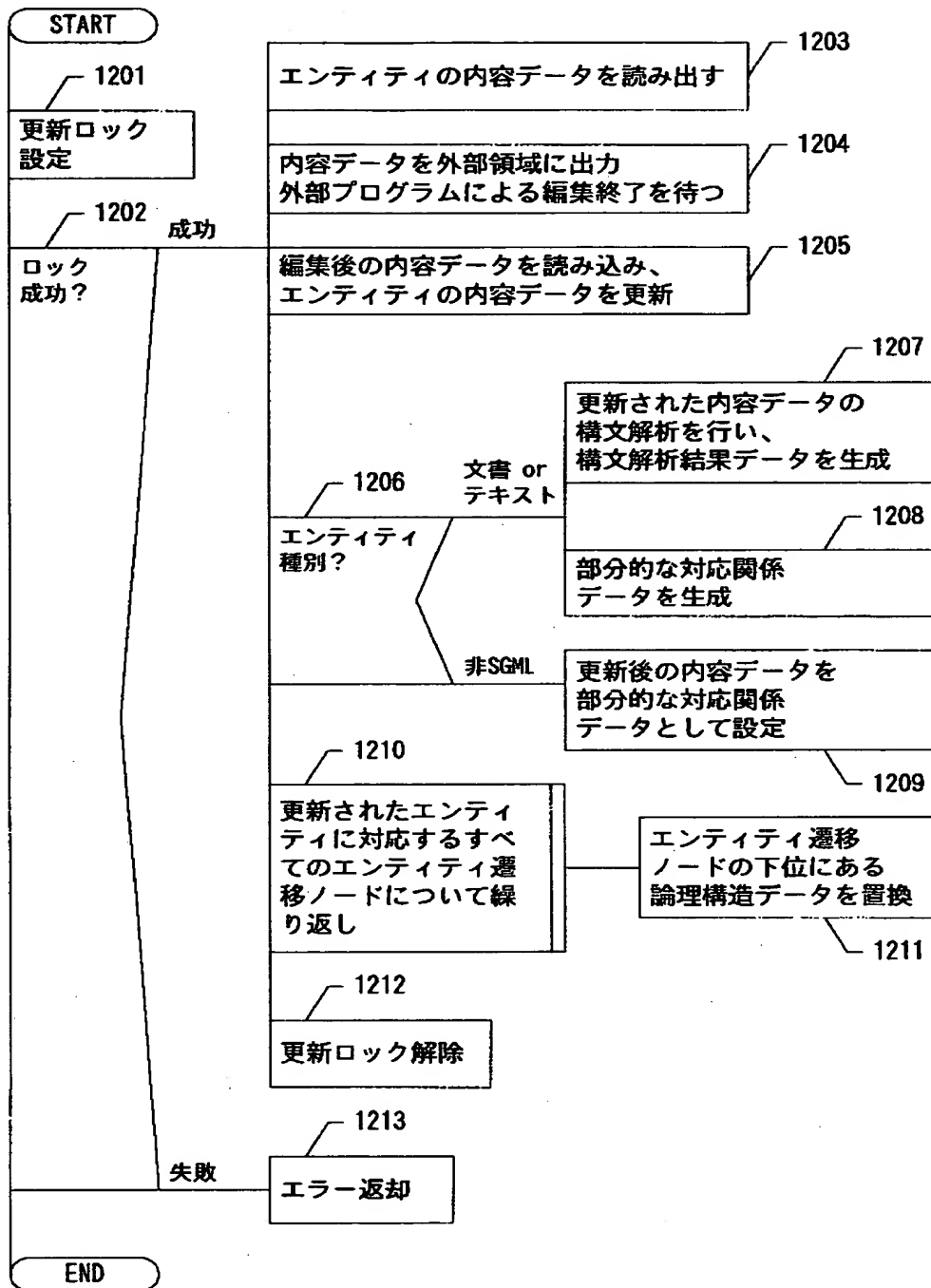
【図 9】

図 9

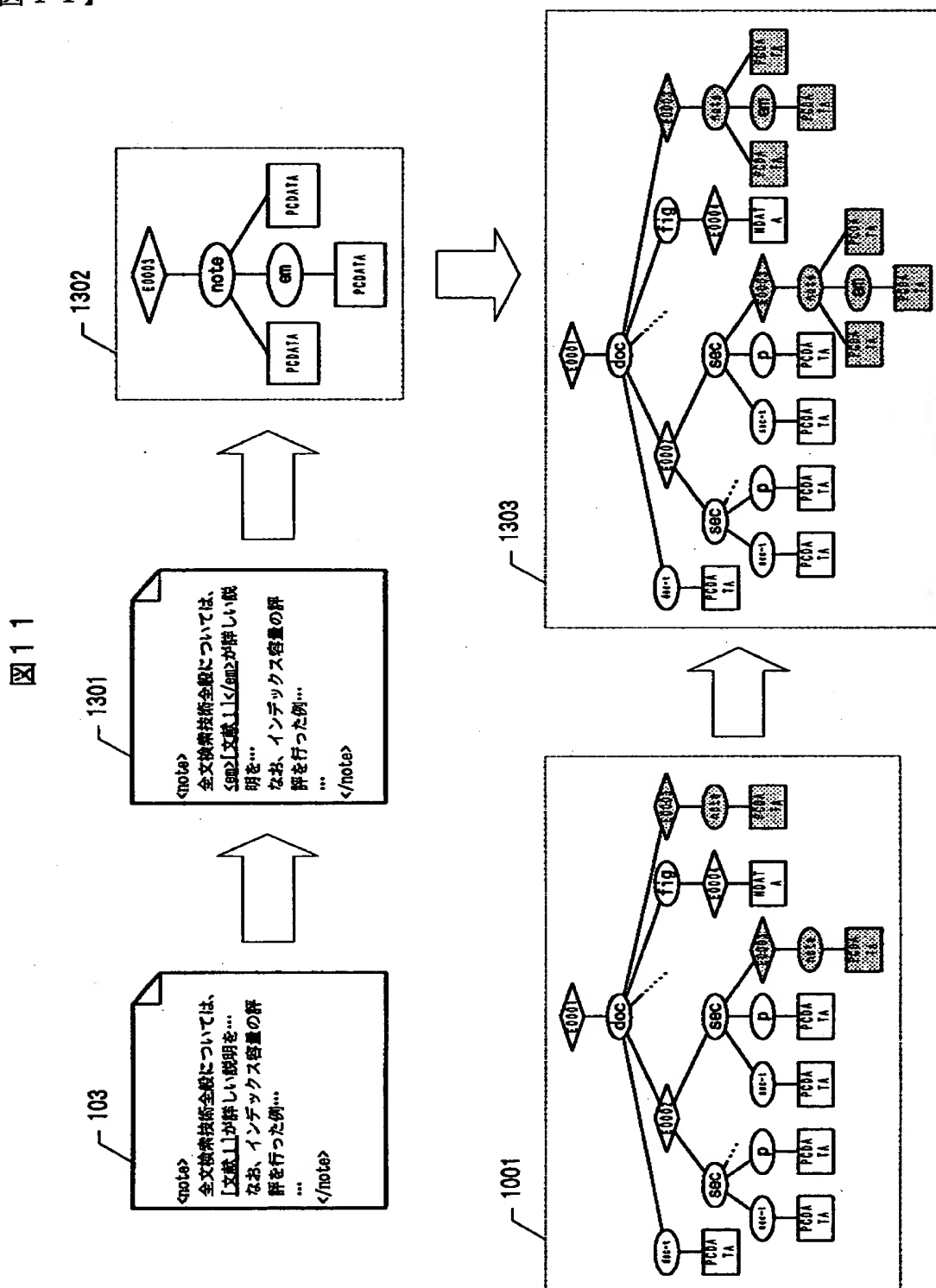


【図 10】

図 10

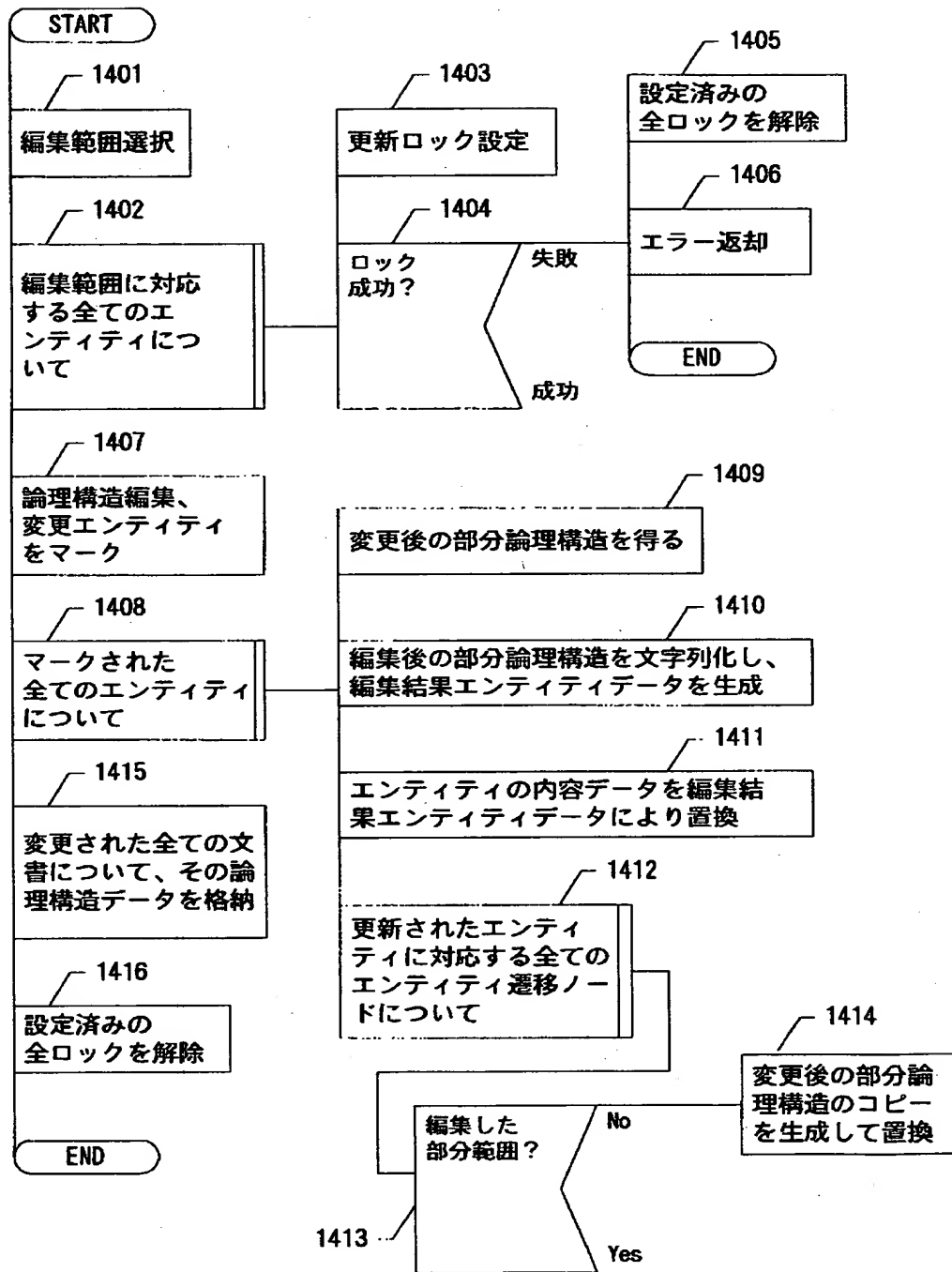


【图 1 1】

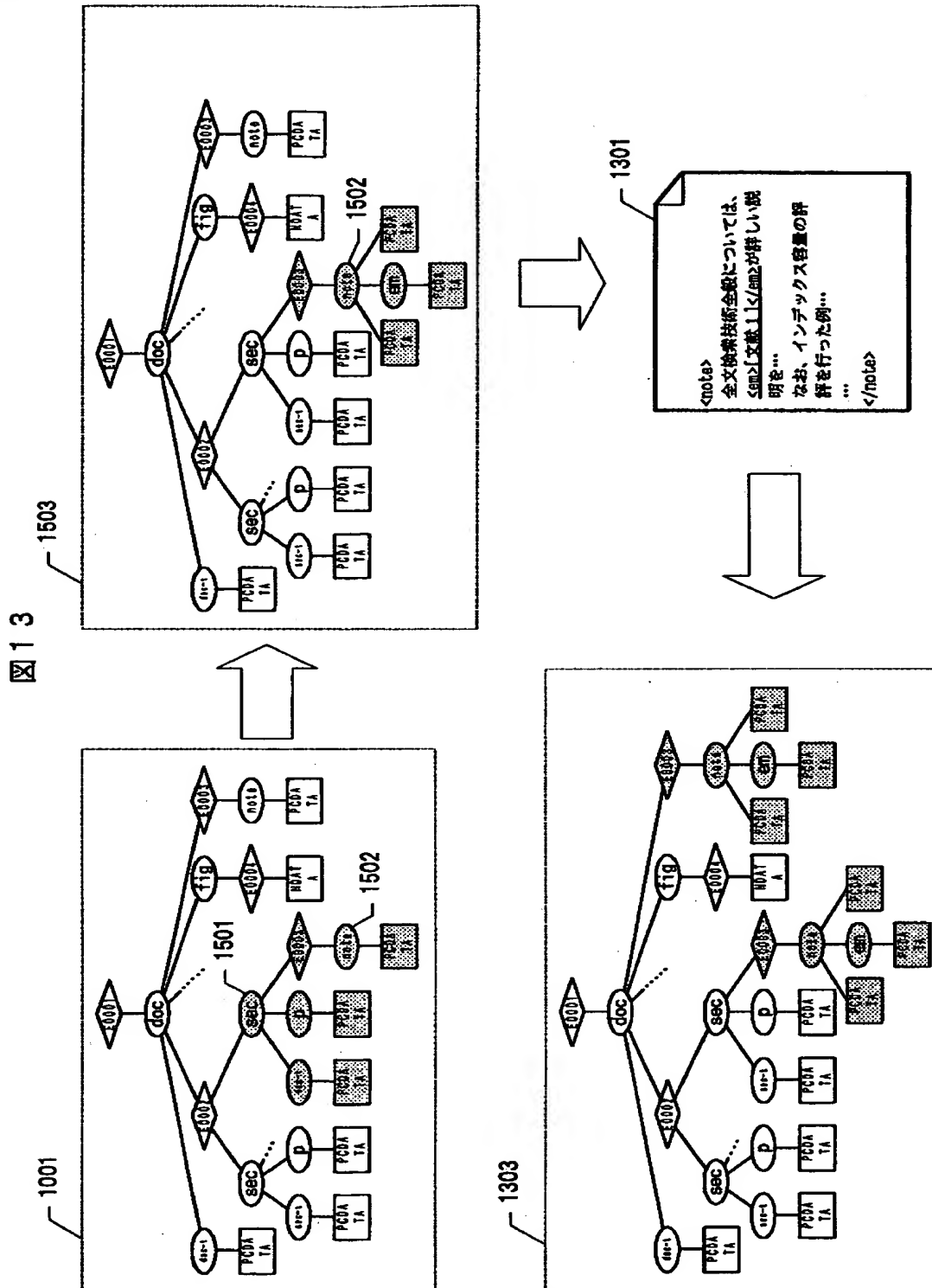


【図 12】

図 12

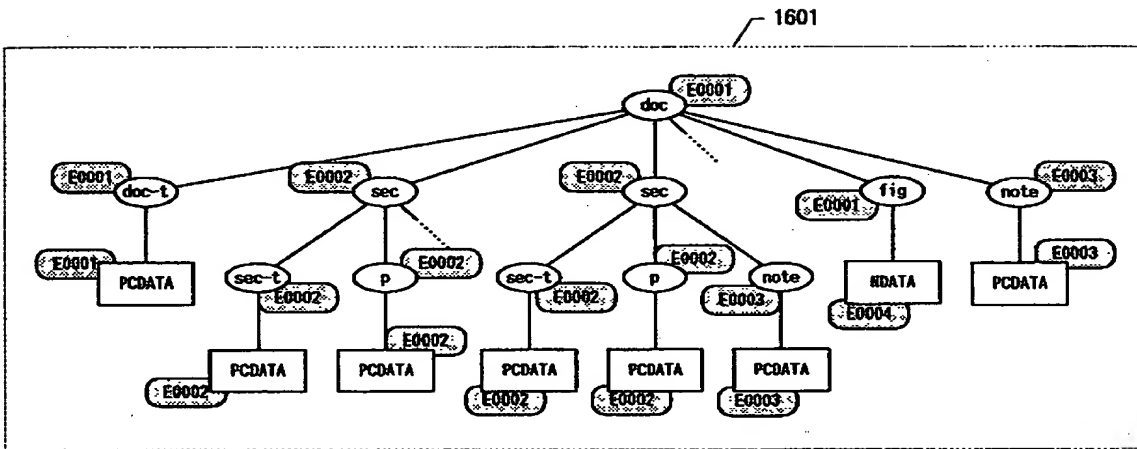


【图 13】



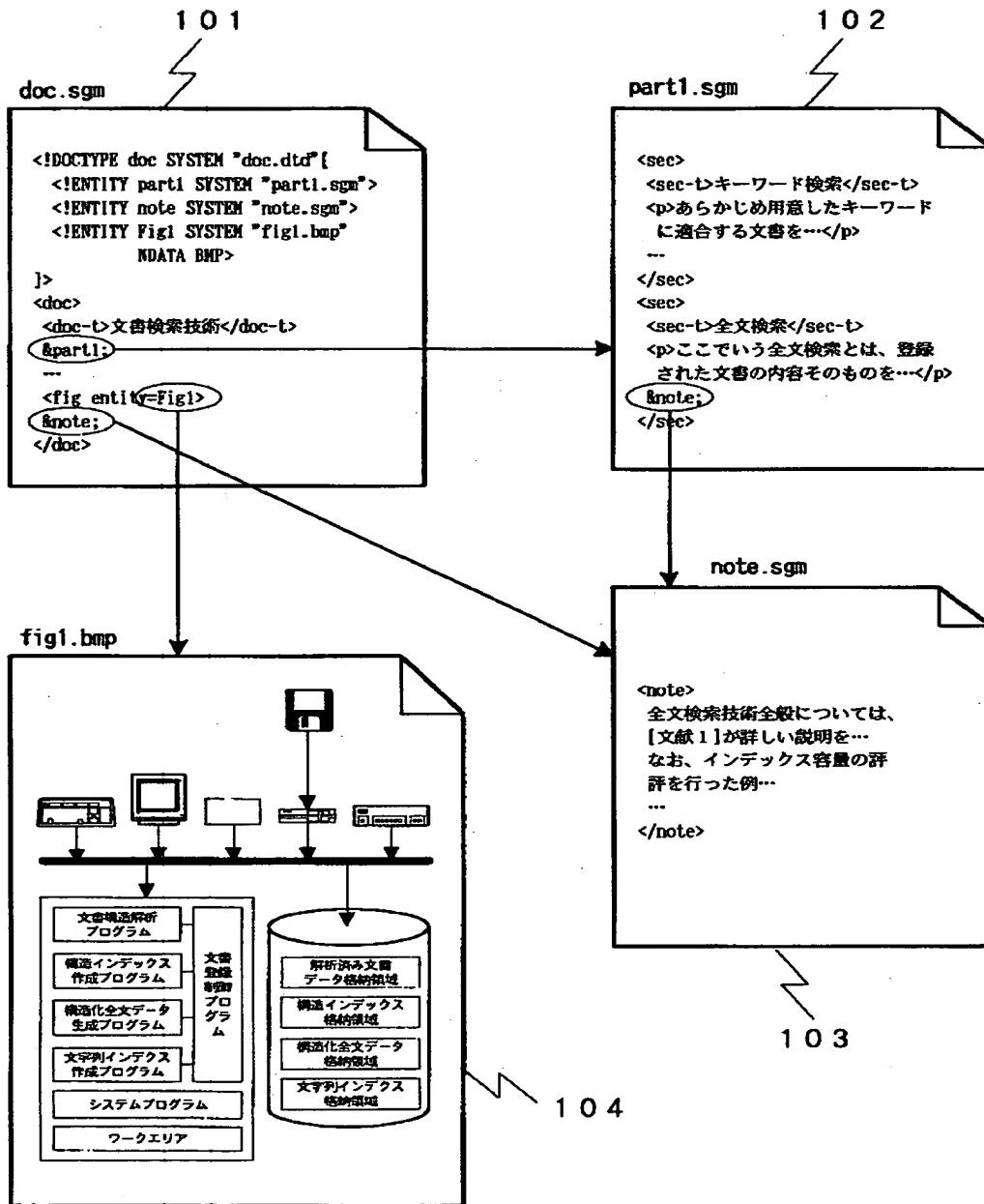
【図 14】

図 14



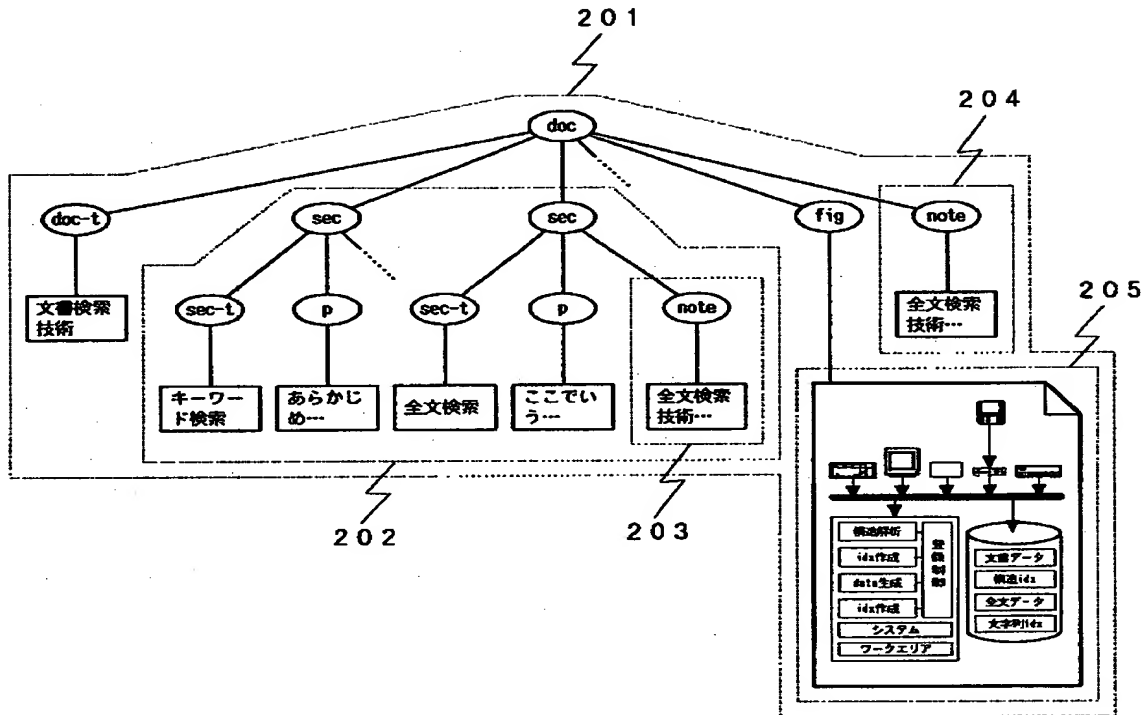
【図 15】

図 15



【図16】

図16



【書類名】 要約書

【要約】

【課題】 登録文書の持つ論理構造と実体構造とを相互に変換可能な形で対応付けて管理することが可能な技術を提供する。

【解決手段】 文書記述言語を用いて記述された構造化文書を管理する構造化文書管理方法において、登録済み文書の実体構造及び論理構造を示す対応関係データから登録済み文書を構成する任意のファイルを更新対象として選択して更新対象ファイルの内容データを更新する内容データ更新ステップと、更新後の更新対象ファイルの実体構造及び論理構造を示す部分的な対応関係データを生成する部分的対応関係データ生成ステップと、前記部分的な対応関係データによって登録済み文書の対応関係データを更新する対応関係データ更新ステップとを有するものである。

【選択図】 図 1

出 願 人 履 歴 情 報

識別番号 [000005108]

1. 変更年月日 1990年 8月31日
[変更理由] 新規登録
住 所 東京都千代田区神田駿河台4丁目6番地
氏 名 株式会社日立製作所